Protein structures elucidating the post-ribosomal biosynthesis of pyrroloquinoline quinone

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Robert Leslie Evans III

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Carrie M. Wilmot, Advisor

July, 2017

Acknowledgements

Soli Deo Gloria: Thank you to my God and Savior. (Isaiah 6:3, Colossians 1:15-17)

Family: Thank you to the love of my life, Karen, for support, encouragement, and strength. Thank you to my mom and dad, Myra and Robert Evans, for always believing and for teaching me to believe. Thank you to my children and their respective spouses, Beth and Roland, Megan and Taylor, Laura and Thomas, Andrew and Emma, for love, support, and laughter. Thank you to my grandchildren, Blaise, Audrey, Rilla, and Felix, for so much wrestling and hugs and giggling and… perspective.

Friends: Thank you to my lab mates, Chao Li and Matt Jensen, for sharing in this adventure and accepting me as more than a colleague… for being my friends, my brothers. Thank you to Morgan Esler; I could not have done this without your help, and I am proud of you. Thank you to the Wackett Lab for consistently generous assistance whenever needed, and especially to lunch gang think tank leaders, James "What's a Creamsicle" Christenson and Lab Mom, Kelly Aukema. Thank you to the Gortner prayer team: James, Matt, Diego, Ben, Bryan, Adam, Rebecca, Anna, and Peter, I could always lean on you. James, you especially, were an answer to prayer (Proverbs 27:17). Thank you to the entire Gortner community from the front office (Lisa!) to the mass spec gang to MicrobTech to Mo Quinn and her wonderful way with miscroscopy [sic].

Colleagues: Thank you to my professional colleagues, mentors, co-conspirators, undergraduate researchers, and high school student volunteers, Peder C., Erik Y., Val K., Chao L., Matt J., Tanner C., Clay E., David S., Morgan E., Andrew W., Ryan M., Jennifer B., Kayla B., Jacob S., Adam L., Anthony B., Gabe B., Adam M., Dustin N., Noah F., Josh A., Jack K... …I must have forgotten some; I had so much great help.

Collaborators: Thank you to the Klinman Lab, U.C. Berkeley (especially Judith Klinman, Ian Barr, and John Latham). John, of course, is now his own bear at the University of Denver. Youlin Xia, thank you so much for your patient tutelage and NMR project guidance.

Institution: Thank you to my dissertation committee members, Sharon Murphy (Chair), Larry Wackett, Hideki Aihara, Hinh Ly, and Carrie Wilmot (Advisor/PI), and the University of Minnesota, college (CBS), department (BMBB), and biotech. institute (BTI).

Mentor: Carrie Wilmot, thank you for taking me into your lab and for teaching me to think critically. You have changed my life. I am humbled by your deep dedication to both your students and your research; I will try to follow in your footsteps.

On a less serious note: Energy in the form of electrons in higher energy bond configurations used for the production of ATP and similar metabolic 'currency' molecules, along with other source materials used in biotic stasis, were often obtained from the following providers. These were essential to my continued work, morale, and contentment: The Original Malt Shop (aka Snuffy's) (Larpenteur Ave.), Mac's Fish & Chips (Larpenteur Ave.), Culver's (all over the place), Dunkin' Donuts (near the synchrotron), and Punch Pizza (Washington Ave.)!

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

About the Enzyme Cofactor, Pyrroloquinoline Quinone

Pyrroloquinoline quinone (PQQ) is a relatively small (330 Da) tricyclic *o*-quinone that has garnered a great deal of research attention since its interactivity with a bacterial aldose sugar dehydrogenase was first reported in 1964.[1] Since the discovery of PQQ, the list of "intrigues and intricacies" surrounding the prokaryotic and eukaryotic biochemistries associated with the molecule has continued to grow.[2]

Early on, the recognition and classification of PQQ as a proteinaceous quinone enzyme cofactor was delayed as it was, and still remains, the only known (poly)peptide-derived quinone cofactor that is not formed *in situ* through post-translational modification of the enzyme polypeptide.[2, 3] Instead, it is non-covalently and reversibly bound to its various cognate dehydrogenase enzymes, and is synthesized separately from a ribosomally synthesized peptide.[2-6] In all other, and more recently discovered, proteinaceous quinone cofactors (e.g. trihydroxyphenylalanine quinone (TPQ)[7] reported in the early 1990s, tryptophan tryptophylquinone (TTQ)[8] also reported early in the 1990s, lysyl tyrosine quinone (LTQ)[9] reported in 1996, and cysteine tryptophylquinone (CTQ)[10] reported in 2001), the cofactor is the result of post-translational modifications to the amino acids of the enzyme active site, which is either self-catalyzed or requires additional biosynthesis enzymes (Figure 1).[2]

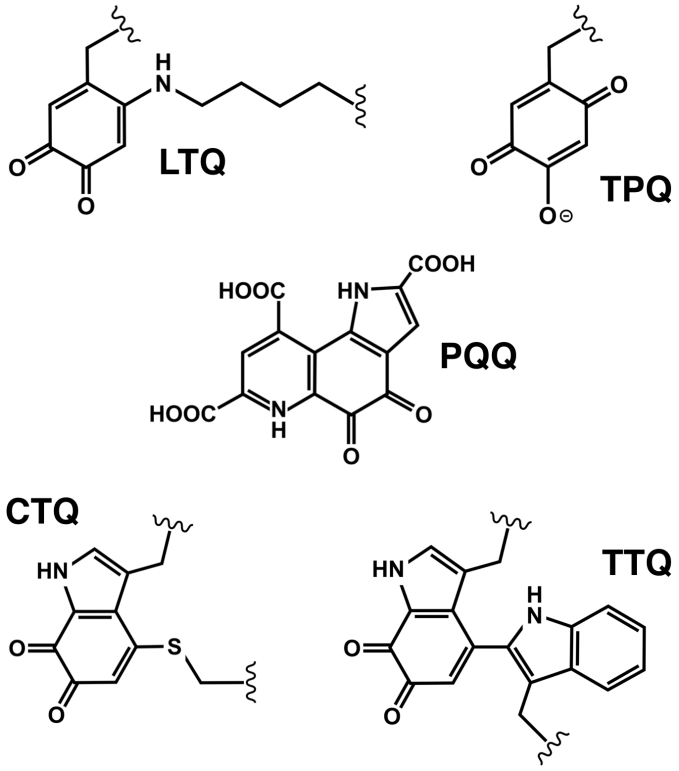

Figure 1. PQQ along with other quinone cofactors.

Regarding PQQ, because it is now known to share significant qualities with other ribosomally produced and post-translationally modified peptides (RiPPs),[11] classification as a unique quinone cofactor may have been premature; perhaps it is better classified, not so much as a proteinaceous quinone cofactor, but as a quinone RiPP.

Gram-negative bacteria represent the majority of species that synthesize PQQ; among these is *Klebsiella pneumoniae*, infamous among human pathogens as responsible for nosocomial infections and the first known pan-resistant (resistant or non-susceptible to all available antimicrobial drugs) bacterial pathogen.[12, 13] Some bacteria that do not make PQQ encode PQQ-dependent aldose sugar/alcohol dehydrogenases in their genomes,

which are expressed under specific nutrient conditions and provide a growth advantage, thus making PQQ a bacterial vitamin.[14-17]

Very recently, data indicated that a eukaryote mushroom, *Coprinopsis cinerea*, may also require PQQ as a coenzyme for a sugar oxidoreductase, even though no PQQ biosynthetic operon (*pqq* operon) exists in eukaryotes.[18] More intriguing is the fact that dietary deprivation in higher eukaryotes, like mice and rats, leads to fertility dysfunction and neonatal growth retardation.[19-22] Similar experiments also linked PQQ-eliminating diets in rats to memory and cognition problems.[23]

The biochemistry behind the probiotic roles of PQQ in higher organisms is not as understood as are its roles in bacterial dehydrogenases. It is known that PQQ is a potent antioxidant (100 times more efficient than ascorbic acid when measured by the number of redox cycles accomplished on a molar basis), and this ability is strongly linked to increases in mitochondriogenesis and mitochondrial health.[24-33] In bacterial pathways, the electrons liberated from the PQQ ⇄ $PQQH_2$ redox cycle are used to generate ATP in the constituent electron transport chains of the organisms.

Natural Products and their Differing Biosyntheses

While the term 'RiPP pathway' is still seen in literature, pathways that produce RiPPs are now referred to as post-ribosomal peptide synthesis pathways (PRPS), because they tailor a precursor peptide that was first synthesized by the ribosomal machinery of the cell.[11]

Compared to other prevalent natural product pathway classifications (e.g. terpenoid biosynthetic pathways, alkaloid biosynthetic pathways, polyketide biosynthetic pathways,

and non-ribosomal biosynthetic pathways), PRPS pathways are relatively simple.[11] Because the bulk of the work is performed by the polymerization activities of the ribosome, the additional enzymes necessary for completing a RiPP are few in comparison to, for example, nonribosomal pathways (NRPs), which require at least one additional enzyme (or catalytic domain) for each peptide residue added. For the purpose of comparison, biosynthesis of the 12-residue, cyclized and branched NRP, bacitracin, requires 12 enzyme modules with a total of 40 catalytic domains.[34] A search of the GenBank database (www.ncbi.nlm.nih.gov/genbank/) shows that only 4 gene products (McjA, B, C, and D) are needed for completion of the 21-residue, cyclized, branched, and knotted RiPP, microcin J25 (Figure 2).

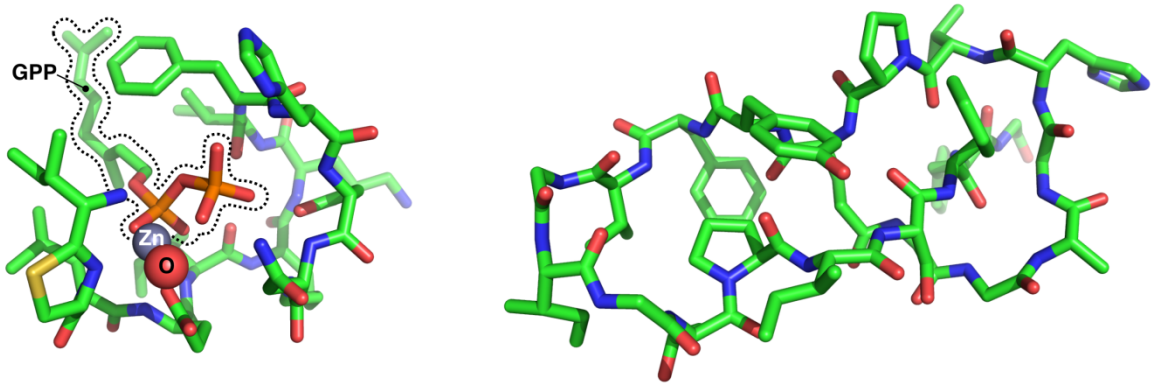

Figure 2. (Left) The antibiotic *NRP*, bacitracin, with Zn-coordinated geranyl pyrophosphate (GPP) outlined with a dashed contour (PDB 1PP5) compared to (Right) the RiPP, microcin J25 (PDB 1PP5).

While all of these classes of natural product biosynthesis offer possible avenues for engineering novel products, RiPPs, while every bit as diverse in size, chemical complexity, and function, are biosynthesized with relative cellular simplicity, fueling an interest in exploring this specific biosynthetic class. Through structural studies, the

Wilmot Lab is contributing to the understanding of the PQQ PRPS, and by extrapolation, the understanding of PRPSs in general.

About PRPSs

RiPPs as well as their respective PRPS pathways are incredibly diverse, and a large collaboration recently standardized RiPP/PRPS nomenclature and classified the RiPPs/PRPSs based on natural product characteristics (Table 1).[11, 35-41] The diverse wealth of natural products listed in this table serves to illustrate the potential this biosynthetic class promises, not only for the richness of products present, but because of the possible novel products that can arise through bioengineering efforts.

Table 1. PRPS classifications based on pathway or RiPP characteristics

Class	Example(s)	Unique Characteristics
Lanthipeptides	Nisin; Subtilin	Lanthionine containing peptides
Linaridins	Cypemycin	Contain thioether crosslinks like lanthipeptides, but are synthesized differently. Might be considered a fifth subclass of lanthipeptides.
Proteusins	Polytheonamides	Once considered the largest NRP, now understood to be a RiPP. An epimerase initiates biosynthesis by producing most, if not all, D-residues contained in the final product. D- and L-residues are in an alternating pattern. Structurally, they contain a cluster of C-methylated amino acids.
Linear azol(in)e-containing peptides (LAPs)	Streptolysin S;	Include azole/azoline rings on non-macrocyclized products.
Cyanobactins	Ulicyclamide; Ulithiacyclamide; Patellamides	N-to-C macrocyclic peptides with proteolytic cleavage and macrocyclization performed by serine proteases.
Thiopeptides	Thiostrepton A; Micrococcin P_1	Macrocycle contains a single piperidine, dehydropiperidine, or pyridine, and several thiazole rings.
Bottromycins	Bottromycin A_2	Include a decarboxylated C-terminal thiazole and a macrocyclic amidine. Similar to proteusins, contain C-methylated amino acids in a series.
Microcins	Microcins B17, C, J25;	With the exception of microcin C, which posses no leader peptide, microcins are tailored with both the leader and core sequences intact. Maturation occurs when the fully-tailored core sequence is cleaved from the leader sequence.
Lasso peptides	Siamycin I; Siamycin II; Microcin J25	Structures contain a specific, knotted, "lasso fold" making them very resistant to denaturing agents and proteases.
Microviridins	Microviridin B; Marinostatin 1-12	Cyclic N-acetylated tri- and tetradecapeptides containing ω-amide and/or ω-ester bonds. Most contain lactams, and all contain lactone linkages, resulting in their tricyclic structures.
Sactipeptides	Subtilosin A; Thurinsin H; Thuricin CD (α & β)	Contain α-carbon to cysteine sulfur (on different amino acids) linkages.
Bacterial head-to-tail cyclized peptides	Cyclic bacteriocins; Enterocin AS-48 is a 70-mer	N-to-C terminal cyclized peptides distinguished by their large size and biosynthetic machinery.
Amatoxins	α-Amanitin	N-to-C cyclized 7-mers containing a tryptathionine crosslink (Trp-to-Cys). Toxic mechanism is different than that of phallotoxins (RNA Pol II inhibition.
Phallotoxins	Phalloidin	N-to-C cyclized 8-mers containing a tryptathionine crosslink (Trp-to-Cys). Toxic mechanism is different than that of amatoxins (binds actin).
Cyclotides	Kalata B1	N-to-C cyclized peptides (from plants) with a cyclic cystine knot (CCK) motif formed of 3 conserved disulfide bonds.
Orbitides	Segetalin A; Segetalin D	N-to-C cyclized peptides (from plants) without disulfide bonds.
Conopeptides and other toxoglossans	Conkunitzins; Conopressins	Peptide venoms from cone snails contain a significantly higher density of disulfide crosslinks and post-translational modifications than other animal venom toxins (e.g. snake or scorpion).
Glycocins	Sublancin 168; Glycocin F	Antimicrobials produced by bacteria that include glycosylation moieties.
A catch-all class for auto-inducing peptides (AIPs), ComX, methanobactin, and *N*-formylated peptides	Methanobactin	For brevity, full descriptions of these are not provided. See the Arnison et al. review for more information.
Catch-all for small RiPPs	PQQ; Pantocin	Smaller molecules not classified above

While the produced natural products are very diverse, there are significant commonalities between the PRPS classifications, the four most significant of these commonalities follow:

1) All RiPPs begin as a ribosomally synthesized precursor peptide.[11]

2) Precursor peptides are most often bipartite (with an N-terminal, leader portion used for recognition, and a C-terminal, core portion providing the amino acids that will eventually comprise the completed RiPP).[11]

3) The tailoring machinery required to complete biosynthesis is comparatively simple.[11]

4) PRPS pathway gene products often include a peptide-recognizing and presenting protein fold, the RiPP recognition element (RRE).[42]

There is much diversity in the RiPP precursor peptides. In some (e.g. PqqA in the PQQ PRPS), only a small number of the residues (two) out of the entire core sequence end up in the final natural product, while in others (McjA in the microcin J25 PRPS), the entire core sequence is included in the final product.[43] The length of these peptides also varies greatly.[11] In some precursor peptides, RRE recognition and presentation require only a leader portion (e.g. PQQ), while in others, a C-terminal recognition sequence is also required (e.g. cyanobactins, amatoxins, cyclotides, and orbitides).[11]

RREs occur in the majority of PRPS pathways.[11, 42] Most often, the RRE is N-terminally attached to the first tailoring enzyme, but occasionally, as is the case with PqqD and MtfB of mycofactocin biosynthesis, the RRE is a small, independent scaffolding protein (Figure 3).[42, 44]

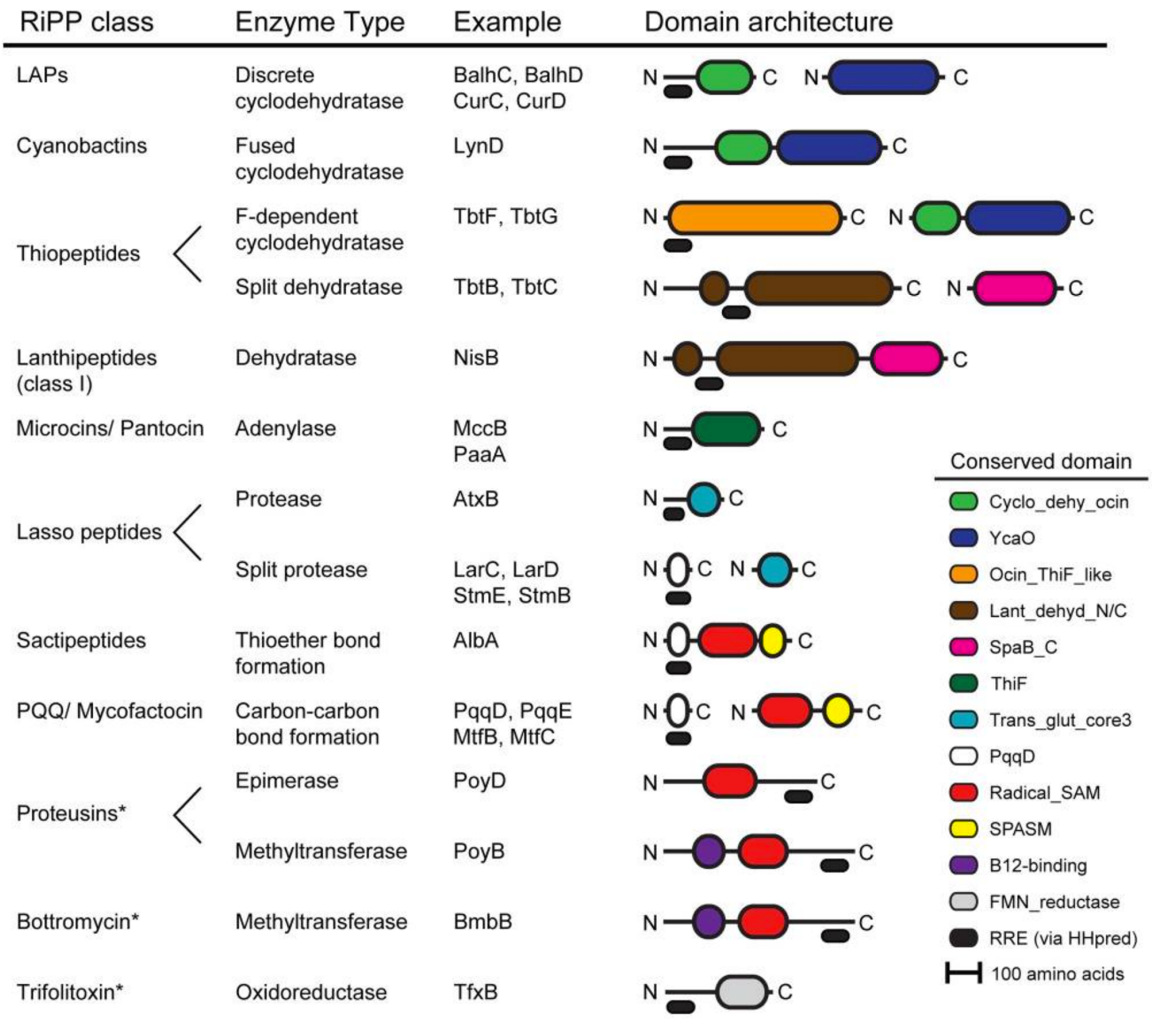

Figure 3. In this figure from Burkhart, et al., 2015,[42] protein domains are shown to scale. The eleven RiPP classes shown represent a subset of the classes mentioned in the Arnison, et al., 2013 review that include RREs. Only 3 out of 11 of the represented classes have C-terminal RREs, and the RRE is separate from the first enzyme in only 2 out of 11.

The PQQ Pathway in Particular

Studies involving the biosynthesis of PQQ are ongoing, and some of the enzyme functions within the pathway are better defined than others, but it is known that five gene products from the *pqq* operon are required for PQQ production.[2]

It is also well understood that all of the carbons and nitrogens of PQQ come from two absolutely conserved residues (Glu and Tyr) found in the C-terminal core portion of the precursor peptide, PqqA (Figure 4).[2]

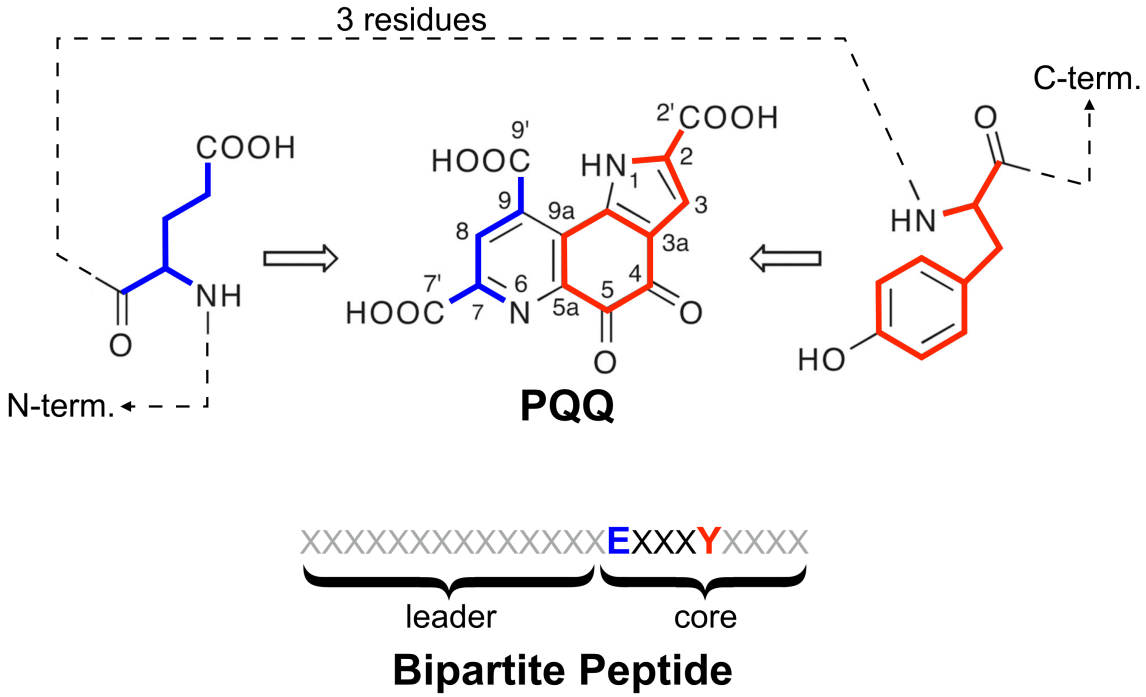

Figure 4. Carbons and nitrogens in PQQ derive from a Glu and Tyr in the bipartite, precursor peptide.

PqqD recognizes, binds, and presents PqqA to the first enzyme of the pathway, PqqE. PqqD is a separate protein in most PQQ pathways (including that of *Klebsiella pneumoniae*), but is C-terminally fused to PqqC in *Methylobacterium extorquens* PqqCD

(MePqqCD). For my dissertation work, I used an artificially truncated PqqD (MePqqD). Later in this discussion I will also mention the PqqC truncation (MePqqC).

PqqE is a member of the radical S-adenosyl methionine (rSAM) enzyme superfamily with a C-terminal SPASM (**s**ubtilosin, **P**QQ, **a**naerobic **s**ulfatase, and **m**ycofactocin) domain.[2] This attached domain is so named for the PRPS pathways that have initiating rSAMs that have also been found to contain the novel domain.[2] PqqE is known to perform Glu-to-Tyr cross-linking.[45] While the chemical structure of the final product of PqqE is not currently defined, the rSAM enzyme most likely produces a molecule that is passed to PqqB for further tailoring.[2] PqqB, then, is thought to produce the substrate for PqqC, 3α-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid (AHQQ).[2] In this dissertation, the product of PqqE, and the unknown substrate for PqqB, will be referred to as pre-AHQQ.

Early *pqqB* knockout studies were inconclusive regarding the requirement for the gene product, PqqB, in PQQ biosynthesis.[2] More recent bioinformatic analyses, however, point to the gene playing an essential role in PQQ production.[2] In unpublished work by our collaborator, Prof. Judith Klinman (University of California, Berkeley), PqqB, a homodimer, appears to be a monocopper-dependent oxygenase, a first for the metallo-β-lactamase superfamily to which it belongs. By sequence identity, the most closely related crystal structure is that of the dimanganese-dependent phosphodiesterase, PhnP.[46] It has already been noted that PqqB had lost one of the conserved residues of the dimetal binding site, and likely contained a monometal active site.[12] This has been confirmed in unpublished work from the Wilmot Lab, but the expected metal ion, Mn^{2+}, bound

weakly, and the most prevalent substitute for manganese, Fe^{2+}, did not bind at all. A PqqB structure was subsequently shown to bind copper with much higher affinity, and the crystal structure solved with copper in the catalytic center. If PqqB is copper-dependent, it would be the first enzyme from that superfamily to use copper.

Early in my graduate work, I was able to trap a putative pre-AHQQ mimic, 5-S-cysteinyl-DOPA (5cysDOPA), in the active site of *Pseudomonas putida* PqqB with a coordinated Zn^{2+} at the catalytic core (Figure 5). Zinc is commonly used as a non-reactive metal substitute in the study of copper-dependent redox enzymes. The 5cysDOPA has a thioether bond in place of the C-C bond expected as the result of PqqE catalysis and contains only two coordinating carboxyl moieties. This allowed the mimic to occupy multiple coordination conformations, which could not be satisfactorily modeled. This ambiguity in metal coordination has led to more recent efforts to trap mimics more similar to the expected structure of pre-AHQQ by another graduate student in the Wilmot Lab, Chao Li. Abbreviated collection and refinement statistics for my unpublished PqqB–5cysDOPA structure are provided (Table 2). Further PqqB work is not included in this dissertation.

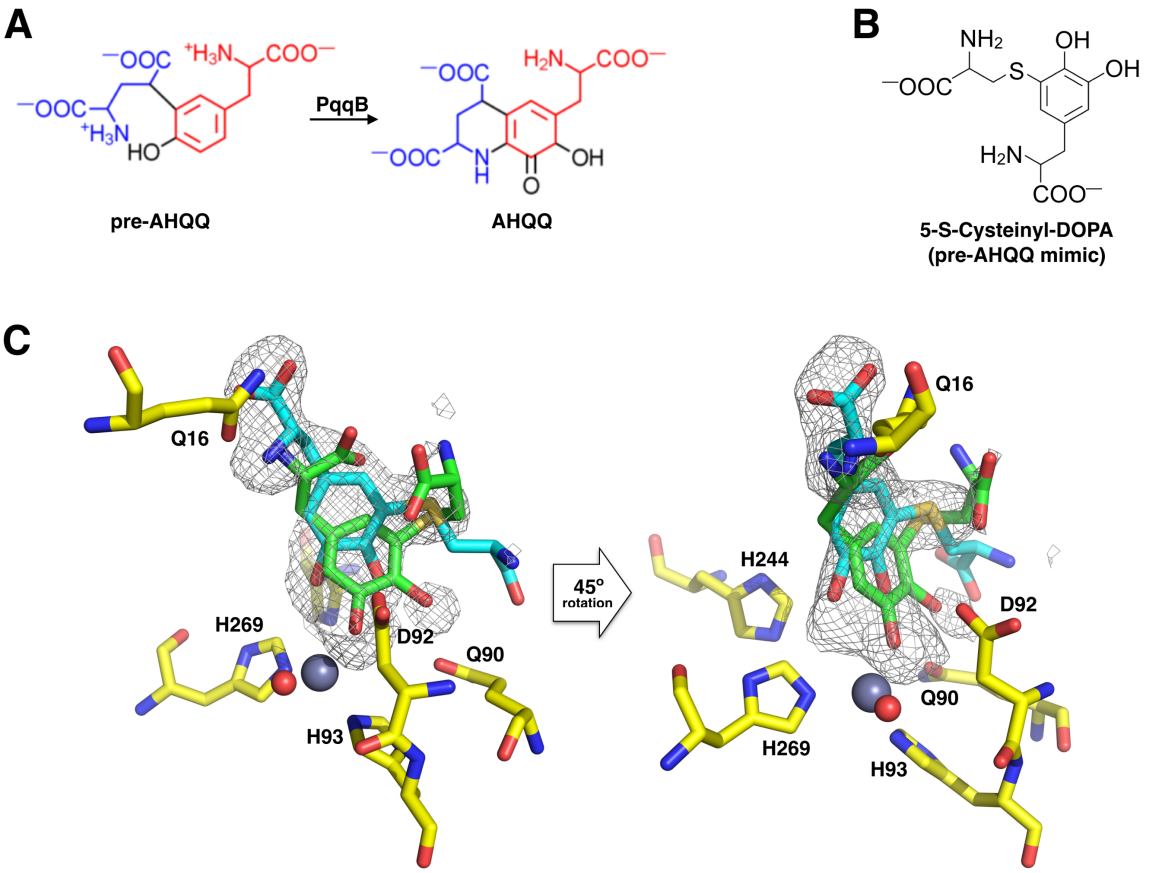

Figure 5. PqqB is thought to convert a pre-AHQQ molecule (putative) to AHQQ (A). The putative pre-AHQQ shown has not been successfully synthesized, therefore 5-S-cysteinyl-DOPA was used as a substrate mimic (B). Two views of the PqqB active site (C). Positive difference density during structural refinement indicates that when crystallized with Zn, PqqB binds 5-S-cysteinyl-DOPA non-specifically. The electron density illustrated here (gray mesh) shows the Fo-Fc map generated by RefMac5 (CCP4 suite). The ray traced density was produced using PyMOL at sigma = +3.0 and carve = 2.5. Two of many possible positions for 5-S-cysteinyl-DOPA are modeled in green and cyan, coordinating residues from PqqB are yellow, zinc is depicted as a gray sphere, the oxygen from a coordinated water is depicted as a red sphere.

Table 2. PqqB–5cysDOPA abbreviated statistics

	PpPqqB
Data collection	
Space group:	$P4_32_12$
R_{merge}:	0.114 (0.829)[a]
$R_{measure}$:	0.119 (0.867)
$I / \sigma (I)$:	18.8 (3.0)
Completeness (%):	99.8 (98.5)
Refinement	
Resolution (Å)	2.35
R_{work}:	0.1778
R_{free}[b]:	0.2377

[a]When used, values in parentheses are for the outer/highest resolution shell. [b]R_{free} is the R factor based on 5% of the data excluded from refinement.

PqqC is the final and best characterized enzyme in the PQQ biosynthetic pathway.[47-49] It is an oxidase that catalyzes an 8-electron cyclization and oxidation of its substrate, AHQQ, to the final product of the pathway, PQQ (Figure 6).[50] It accomplishes catalysis by activating O_2 in the absence of either an external cofactor or a metal, through a progression of PqqC conformational changes, from an open resting enzyme to closed product complex.[48-50] These conformation changes reposition multiple, coordinating residues, including basic amino acids required for proton abstractions (tautomerization steps of B_1, B_2, and B_3 in Figure 6 and panel C in Figure 7).[50] Fully open and fully closed structures for *Klebsiella pneumoniae* PqqC (KpPqqC) have been solved crystallographically (Figure 7).[49, 51] My work on MePqqCD and the truncation, MePqqC, involved finding alternative crystal morphologies for *in crystallo* trapping of catalytic intermediates. The search for alternative crystal morphologies, other than that already published for KpPqqC, began after our failed attempts to crystallize wild type KpPqqC in a form that could bind PQQ and undergo PqqC closure following addition of PQQ to the crystal containing solution failed. The inability to replicate published KpPqqC

crystallization conditions that were reported to support the PqqC structural changes, puzzled us until earlier this year (January 2017), when I noticed that the deposited "wild type" crystal structures (PDB IDs 1OTV and 1OTW) are actually A21D mutants.[49] The Asp21 residue appears to play a critical role in crystal packing, leading to a $P2_12_12$ space group, and explains why our *true* wild type (Ala21) packed as *I*222 and did not bind PQQ during crystal soaking experiments. Additionally, unpublished soak experiments by a former Wilmot Lab postdoc, Val Klema, using crystals of two mutants of KpPqqC (R179S and Y175F), demonstrated binding to PQQ, but both failed to adopt the fully closed conformation of the PqqC product complex (Figure 8). KpPqqC Y175F has been demonstrated to adopt a fully closed conformation when co-crystallized with PQQ, but R179S is known to lead to a more open conformer.[52] However, KpPqqC Ala21 is a surface residue not involved in the active site, and so efforts are now ongoing to crystallize the A21D KpPqqC mutant to reproduce the crystal form suitable for trapping catalytic intermediates “on-the-fly” in PqqC crystals.

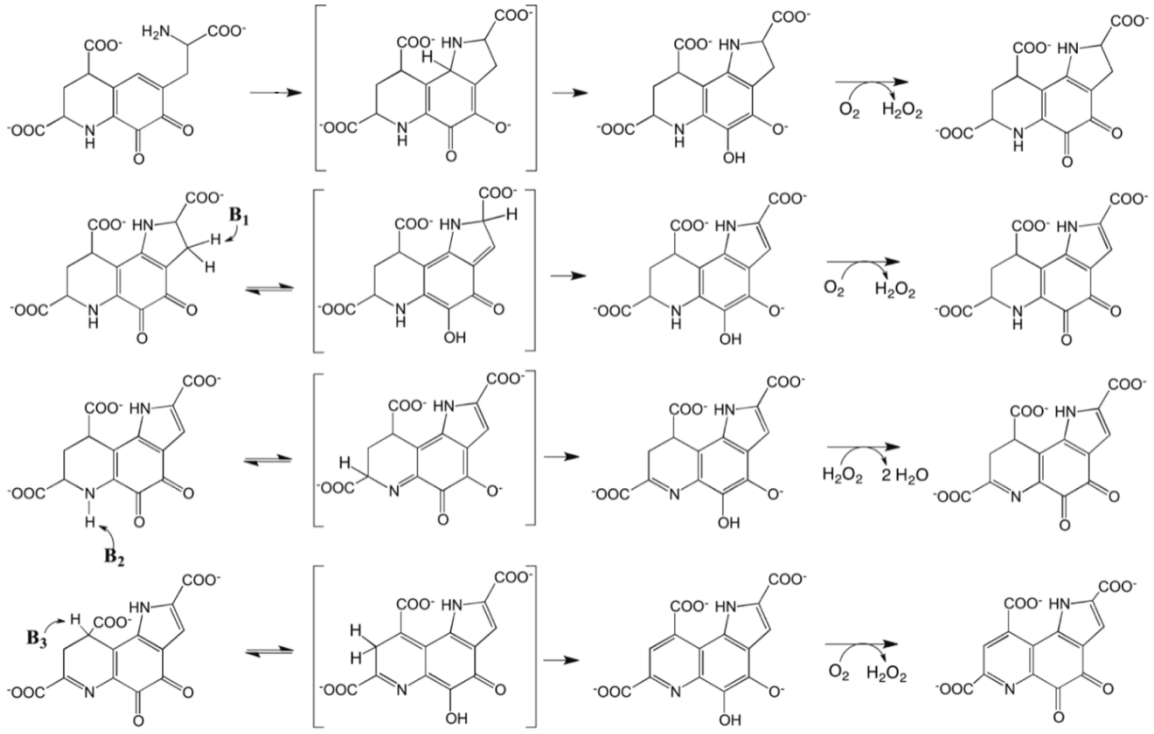

Figure 6. Proposed PqqC mechanism by our collaborators in the Klinman Lab, U.C. Berkeley. AHQQ is represented at the top-left, and PQQ at the bottom-right. Each row represents one of four 2-electron oxidation steps that correspond to progressive closure of the enzyme. This figure was adapted from scheme 1 in the paper by Bonnot, et al., 2013.[50]

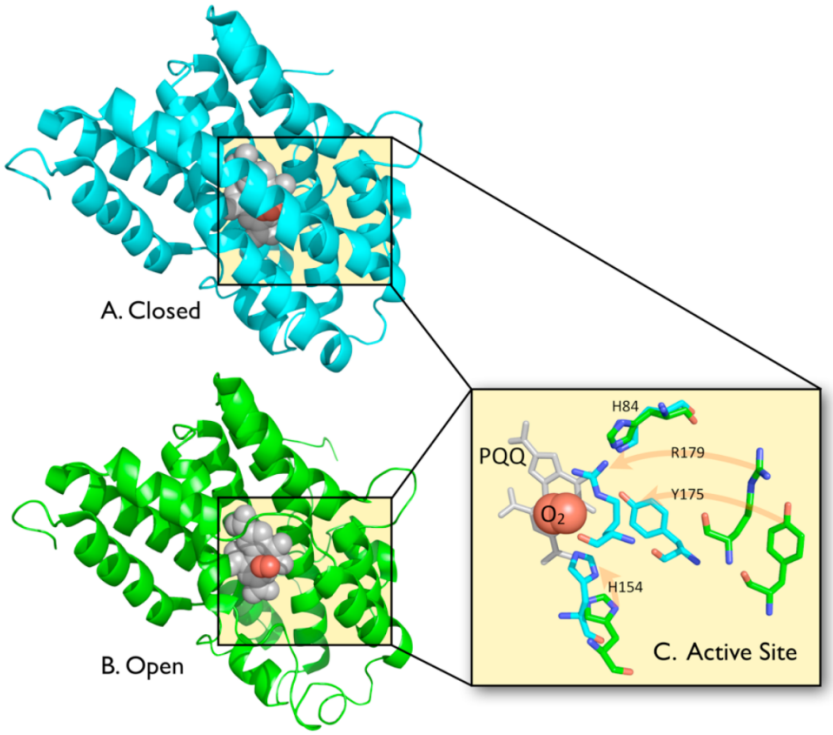

Figure 7. KpPqqC in its open (A) and closed (B) conformations, corresponding to resting enzyme and product complexes, respectively, along with some of the largest movements of active site residues (C) when going from open (green) to closed (cyan).

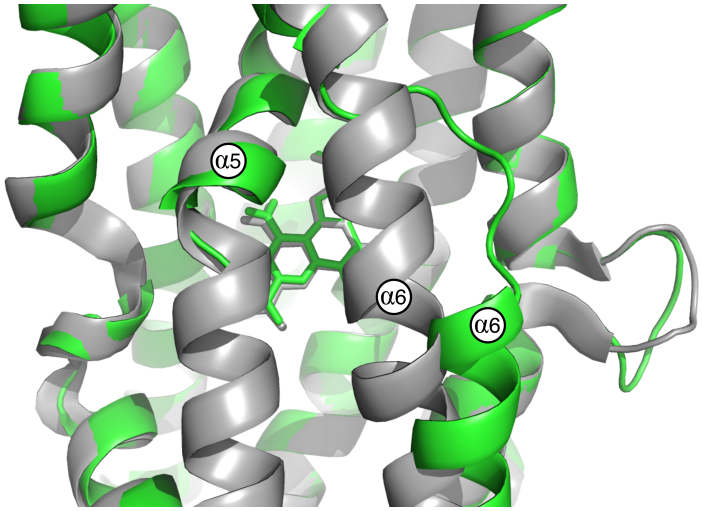

Figure 8. Overlay of KpPqqC *Y175F* soaked in a PQQ-containing solution (green) and the KpPqqC Y175F-PQQ co-crystallization complex (grey) of Puehringer et al (2010) (PDB ID 3HLX).[52] Helices 5 and 6 remain disordered in the PQQ soak experiments, but close in the co-crystallization.

The putative protease, PqqF, while not always present in the *pqq* operon, appears to play a role that can be performed by other proteases, as PQQ biosynthesis still occurs with no phenotypic consequences when the *pqqF* gene is knocked out in organisms that contain it.[2, 18]

In summary, the PQQ biosynthetic pathway begins when PqqD recognizes and presents the ribosomally-produced peptide, PqqA, to the first tailoring enzyme, PqqE. Where the putative protease, PqqF, acts in the pathway is not known, and it could conceivably follow PqqE, excising the cross-linked Glu-Tyr from the PqqA peptide, or it could follow PqqB, a putative copper-dependent oxygenase. Although these steps still need to be defined, the result is AHQQ, which is finally oxidized to PQQ by PqqC.

Research significance and specific aims

Because of its predominant production by Gram-negative pathogens, the PQQ biosynthetic pathway was initially considered a possible target for novel antibiotic development. More recent scientific interest has centered on PQQ biosynthesis as a model pathway for PRPS natural products. The specific aims of my research have focused on the *initiating* and *final* steps of PQQ biosynthesis. Initiation involves the RRE, PqqD, binding the precursor peptide, PqqA, for presentation to the first tailoring enzyme, PqqE. The RRE-precursor peptide interaction is found in the majority of PRPS pathways, and the natural separation of RRE from enzyme encountered in the PQQ and mycofactocin pathways makes it ideal for structural studies by nuclear magnetic

resonance (NMR).[44, 45, 53] The final step in the pathway is the conversion of the intermediate, AHQQ, to the final natural product, PQQ, catalyzed by PqqC.

The initiating steps of PQQ biosynthesis provide novel opportunities to explore emerging biochemistry. rSAM-SPASM enzymes are widespread, with ~14000 annotations in the Interpro sequence database (https://www.ebi.ac.uk/interpro/) and these are often associated with PRPSs. The prevalence of rSAM enzymes in RiPP biosynthesis is very likely linked to the ability of radicals to initiate difficult chemistries such as oxidative decarboxylation,[53, 54] thioether bond formation,[55-60] and the C-C bond formation required for Glu–Tyr crosslinking in PQQ biosynthesis.[2, 45, 61] Bioinformatics suggests that ~50% of all PRPS rSAM-SPASM enzymes have an identifiable N-terminal RRE domain as part of the polypeptide, although in PQQ, the RRE is present as a separate protein. The final RiPP products from PRPSs that feature rSAM-SPASM enzymes are chemically diverse with a wide range of physiological functions: antibiotic (thurincin H and subtilosin A), growth regulation (sporulation killing factor), signaling (streptide), and redox cofactors (PQQ).[11]

Unfortunately, the only crystal structure of a rSAM-SPASM enzyme is that of anSME, which is not part of a RiPP pathway but activates a sulfatase under anaerobic conditions through post-translational modification.[62] Therefore, the structural details of how peptide substrates are presented to the prevalent rSAM-SPASM enzymes found within diverse RiPP biosynthesis pathways are currently unknown. The results presented in chapters 2 and 3 of this dissertation provide crucial insights into RRE interactions between precursor peptides and the prevalent rSAM-SPASM enzymes of RiPP pathways,

while that of chapter 4 focuses on crystal structures of MePqqCD and MePqqC as steps towards trapping catalytic intermediates in PqqC crystals.

Three aims of my research

1. Solving the structure of PqqD. While a non-physiological, domain-swapped dimeric form of *Xanthomonas campestris* PqqD (XcPqqD) exists in the PDB, no monomeric (physiological) structure existed.[63] A structure of MePqqD would not only be the first physiological structure for PqqD proteins in general, but also the first structure of a stand-alone RRE across all PRPS pathways.

2. Identify binding surface residues on PqqD. I desired to learn how PqqD binds to PqqA, and how PqqD + PqqA together bind to PqqE. The majority of PRPSs include RREs with a fold very similar to that of PqqD, but only a few of these are independent/stand-alone RREs; most RREs are domains of the first tailoring enzyme. I wanted to compare this novel, independent RRE to known/deposited constituent RREs in complex with precursor peptides within the PDB, to learn similarities and differences in peptide binding and presentation in PRPS pathway initiation. Additionally, I wanted to define how the RRE interacted with a rSAM-SPASM enzyme, an interaction where there is currently no structural information.

3. Crystallize and solve structures of MePqqCD and the MePqqC truncation. Multiple structures for KpPqqC have been deposited in the PDB, including bound and unbound forms of wild type (with a A21D surface mutation) and active site mutants. What has not been accomplished is the trapping *in crystallo* of AHQQ-to-PQQ intermediates in the PqqC active site. Structures with trapped intermediates would

elucidate mechanistic details for how PqqC catalyzes the conversion of AHQQ to PQQ. To this end, I turned to MePqqCD and MePqqC.

Methods in Structural Biology

Overview

Whether a protein functions as a structural building block, scaffold for mediating protein-protein or protein-biomolecule interactions, hormone, receptor, reaction-catalyzing enzyme, or any other function(s) from a long list of possibilities, proteins are at the very heart of biochemistry, and a three-dimensional structure elucidating the placement and relationship of the constituent atoms is vital to understanding intra- and interactivities with other biomolecules. An analogy can be drawn between studying protein structures (using techniques like X-ray crystallography (XRC), nuclear magnetic resonance (NMR), and/or cryo-electron microscopy (CEM)) and interacting with an automobile; one can learn to drive a car without taking it apart, but one will never really understand how to fix the car, unless one removes and studies each and every part. Techniques like XRC, NMR, and CEM allow structural biologists to do just that, to extrapolate from the detailed structures of proteins how they work, while gaining deeper understanding of function and molecular interactions. Structural biology with its corresponding biophysical techniques is a crucial part of understanding life, the mechanisms of disease, and even rationally engineering vaccines, drugs, and biotechnological tools from our structural knowledge of proteins.

X-ray Crystallography

Protein XRC is predicated upon obtaining a protein sample of sufficient purity and solubility to allow for screening experiments of multiple variables aimed at discovering the—often very specific—conditions under which that protein will form crystals.[64] These variables include protein construct and concentration, buffer compositions, pH, temperature, precipitant concentrations and compositions, time, and so forth.[64]

The very short wavelengths useful for structural determination diffract well enough to solve structures at resolutions as high as 0.5 Å, but these energies interact weakly with the small amount of matter, specifically electrons, in a single protein. This limitation is overcome by the amplification effect provided by the very large number of protein molecules contained in a protein crystal. Once a growth condition is found to produce protein crystals, the crystals are screened to determine the quality of diffraction (Figure 9). External indicators such as optical clarity, color, size, and shape are not reliable predictors for diffraction quality. While larger crystals contain more protein and therefore greater diffraction amplification than their smaller counterparts, in some morphologies, larger crystals provide more opportunity for diffraction-degrading misalignments and imperfections in the lattice, leading to poor diffraction.[64] In the end, the process of finding specific conditions that lead to suitably-diffracting crystals, can take weeks, months, years, or even fail (Figure 10). Intrinsically disordered proteins by their very nature are unsuitable subjects for XRC as they tend to not form highly-ordered crystals. Cryo-cooling crystals to reduce radiation damage during the experiments can also lead to reduced diffraction quality.

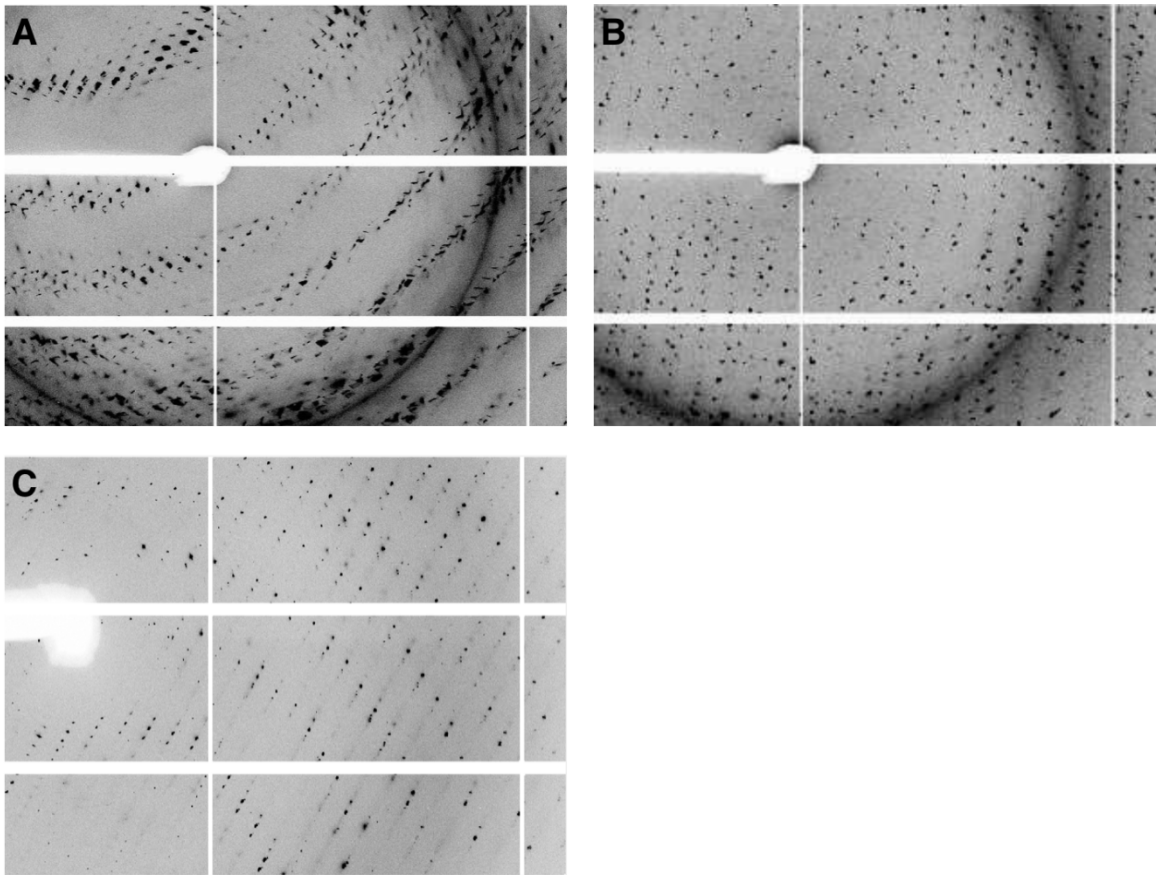

Figure 9. Diffraction quality. Subtle misalignments in crystal packing leading to spot smearing or mosaicity (A) make peak integration difficult or impossible. Multiple crystal lattices superimposed on top of one another (B) can be caused when crystals grow into one another, or could simply mean that multiple crystals were harvested together, accidentally. The portions of the diffraction seen in panels A and B show strong ice rings, which obscure reflections. While not perfect, diffraction patterns with minimal mosaicity or icing, and clear, round reflections (C) can be converted into interpretable electron density. A concept not directly illustrated in this figure is that of resolution. Reflections farther from the center represent higher resolution data (more structural detail), and that is also considered a metric of diffraction quality.

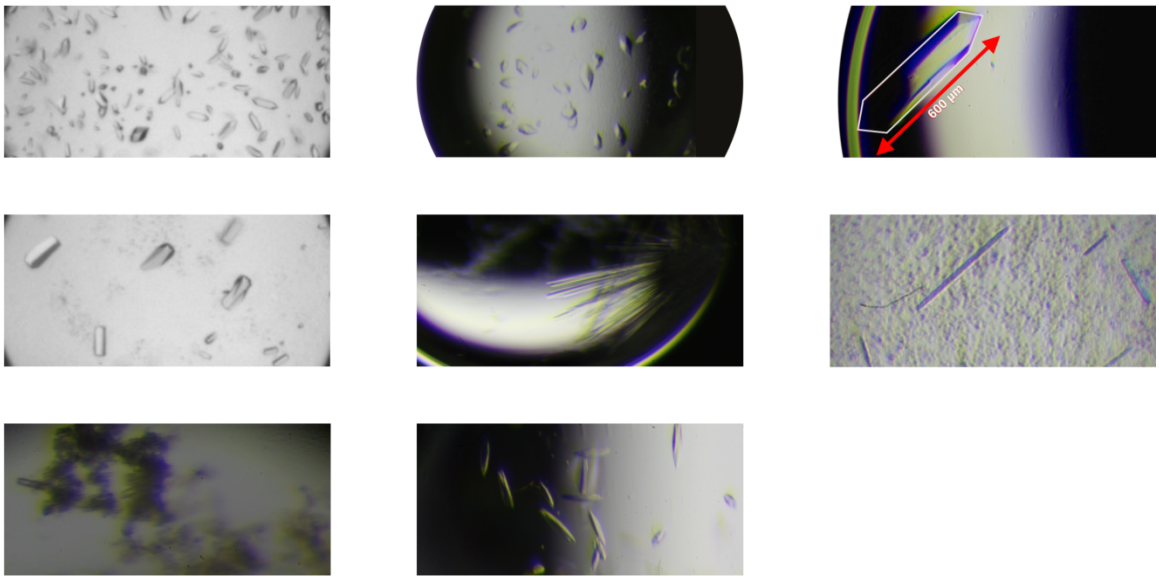

Figure 10. MePqqCD crystal morphologies. After broad screening for crystallization at the UMN Kahlert Structural Biology Laboratory's crystal screening facility, multiple "hits" lead to dozens of variable combinations to pursue. While the crisp, optically-clear, 600 µm crystal in the upper right pane (outlined for clarity) was visible to the naked eye, its diffraction was inferior to smaller crystals of a different morphology (not pictured) that led to the 2.85 Å structure presented in chapter 4 of this dissertation.

Provided that a given protein can be crystallized and sufficient, high quality diffraction data can be collected, another hurdle remains for the crystallographer, that of solving the phases for the thousands of reflections that comprise a dataset. While the detector can collect information about the position of an individual spot (relates to specific *hkl* planes in the crystal lattice) and intensity (relates to the quantity of electron density (matter) in that plane), no detector can measure the phases of the diffracted photons associated with that spot, and yet, that phase information is even more critical to creating an the electron density map than the position and intensity data collected by the detector (Figure 11).[64] This is what is known as the "phase problem" in XRC.[64, 65]

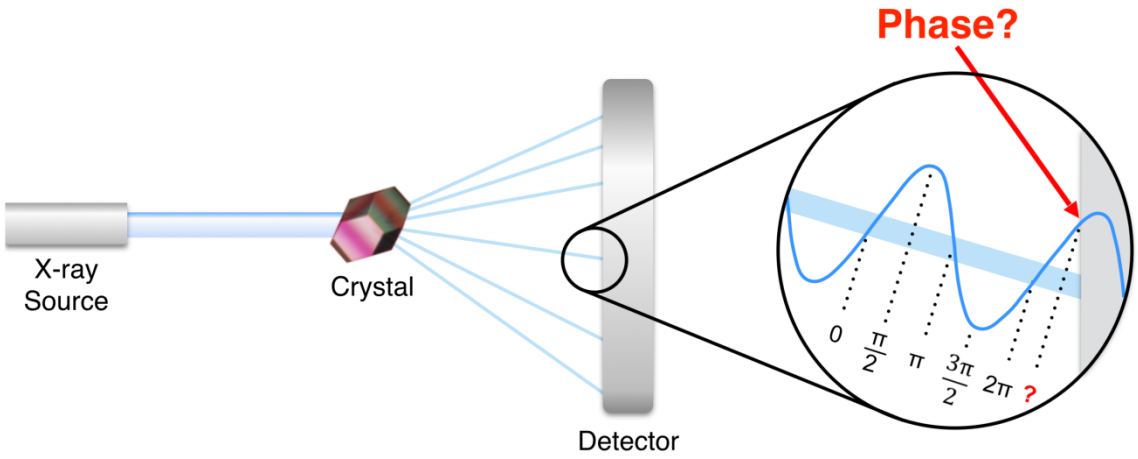

Figure 11. Solid state detectors cannot detect the phases of the diffracted photons. Graphic was co-developed by the author and Matthew Jensen for 2015–2016 presentations given to Dr. Larry Que's Chemical Biology of Enzymes, CHEM 8412, at UMN.

There are many methods that can be employed to solve the phase problem, from more involved, *de novo* methods, such as single- or multi-wavelength anomalous diffraction (SAD or MAD), to relatively simple methods, like molecular replacement, where phases are essentially "borrowed" from a previously-solved, homologous protein structure and then refined to a statistically-guided endpoint.[64-66] Fortunately, for both of the PqqC structures presented in this research, phases from a close homologue were used successfully. Details for the two PqqC structures' initial data quality and refinement statistics are provided in chapter 4.

NMR

Similar to XRC, solution NMR experiments require sufficient concentrations of purified—and usually isotopically-labelled—protein in buffer conditions that do not suppress or interfere with the detection of resonance signals; ionic salts like sodium chloride can suppress signal (Figure 12). For proteins where structural stability requires

buffer conditions with relatively high ionic strength, this can present a challenge. Fortunately, the PqqD in our research remained in solution at concentrations upward of 10 mg/ml without signal degrading salts (e.g. NaCl and KCl) through the use of a simple, non-interfering buffer (25 mM potassium phosphate). As with XRC, some proteins are not suited for solution NMR due to solubility, and the complexity of NMR spectra limit the size of protein that can be studied.

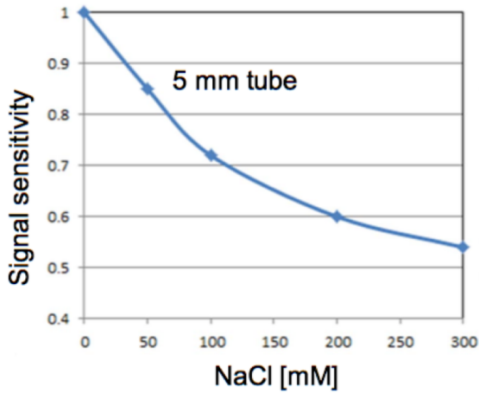

Figure 12. NMR signal sensitivity and therefore, signal-to-noise, diminishes with increasing salt concentrations. Figure from data collected by collaborator, Dr. Youlin Xia, on a Minnesota NMR Center 700 MHz instrument equipped with a 5mm TXI cryoprobe.

When atoms with a nuclear spin quantum number (*m*)—those with an odd number of nucleons (protons and neutrons), such as ^{1}H, ^{13}C, and ^{15}N—are placed in a strong magnetic field (B_o), a couple of things happen; 1) the energy gap between nuclei in the low energy spin state ($m = +1/2$) and those in the high energy spin state ($m = -1/2$) separates and broadens (Figure 13), and 2) the nuclear magnetic spins begin to precess (wobble) at a specific frequency known as the Larmor frequency.[67]

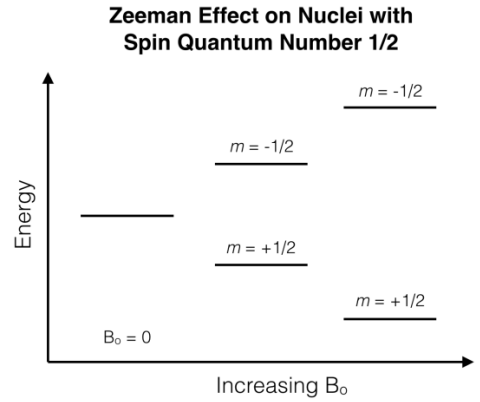

Figure 13. Zeeman Effect. Magnetic field-induced splitting and broadening for the energy states in nuclear spin.

The perturbation of these precessing nuclear magnetic spin resonances lies at the heart of NMR, and the energy-gap splitting and broadening effect, called the Zeeman effect, amplifies the ability to "see" these perturbations, because the stronger the magnetic field, the greater the effect (Figure 13).[67] Broadening of these energy differences is directly tied to the sensitivity of the NMR instrument and peak-resolving power.[67] Stronger magnets (NMR instruments are often referred to as, simply, magnets) provide sharper peaks and are able to resolve spectra where smaller magnets might fail. Larger molecules (i.e. proteins) tumble more slowly in solution than smaller molecules, and this slower tumbling leads to peak broadening.[67] This is why larger NMR instruments

(upward from 650 MHz) are most used for macromolecular studies, while smaller instruments (650 MHz and smaller) may be employed for simpler, organic molecules. Time is also a factor in resolving power, as a large number of repeated spectra collected over a longer period of time, when averaged, improves signal-to-noise, thus sharpening peaks and potentially getting around the need for a bigger magnet. As instrument time is precious, striking a balance between longer repeat experiments on a smaller magnet versus shorter duration experiments on a larger magnet must be weighed against machine availability, cost of operation, and so forth. For most of the experiments presented in this study, which required 2D and 3D spectrum collection, the 850 and 900 MHz instruments at the Minnesota NMR Center were used, but quick (30 minute) $^{1}H,^{15}N$-HSQCs to monitor protein stability/degradation were often performed on a 700 MHz instrument.

NMR experiments consist of very specific pulse sequences consisting of timed bursts of electromagnetic energy at FM-radio frequencies, geometrically induced *orthogonal* to B_o.[67] When matching (resonating with) the Larmor frequency of a specific nuclei (e.g. ^{1}H, ^{13}C, and ^{15}N), the magnetic wave component of the electromagnetic pulse perturbs that nucleus's magnetic spin precession, causing an unnatural number (higher than the number attributed to the normal Boltzmann distribution) of nuclei to experience spin inversion, effectively knocking nuclei to the higher energy (m = -1/2) spin state.[67] The instrument then detects the decay times as the perturbed atoms return to their pre-perturbed, normal distribution energy states (m = +1/2); this is the free induction decay (FID) signature, and it looks like a damped sinusoidal oscillation.[67] Performing a Fourier transform (FT) on an FID effectively converts the spectra from the time domain to the

frequency domain allowing us to observe peaks representing individual atoms or specific atom-to-atom relationships (Figure 14).[67] Once converted to the frequency domain, the frequencies are normalized to 'parts per million' notation (ppm or δ), allowing for spectra from any instrument of any magnetic strength to be compared to any other instrument's spectra.[67]

By combining coordinated NMR pulse sequences into multi-dimensional experiments that take advantage of specific nuclear Larmor resonances in the macromolecule, spectral peaks representing individual atoms can be identified, and 1-, 2-, and 3-dimension positional relationships between atoms can be deduced. Resonance interactions can propagate through bonds or through space.[67] When resonance interactions occur through space, they are called Nuclear Overhauser Effects (NOEs).[67] These positional relationships along with knowledge of the primary sequence of the protein, allow for deductive assignment of peaks and can ultimately lead to structural determination.

In addition to structural studies, NMR can be used to elucidate protein-protein/substrate interactions. By observing peak shifts induced by adding binding partners, residues involved in binding can be identified.[67]

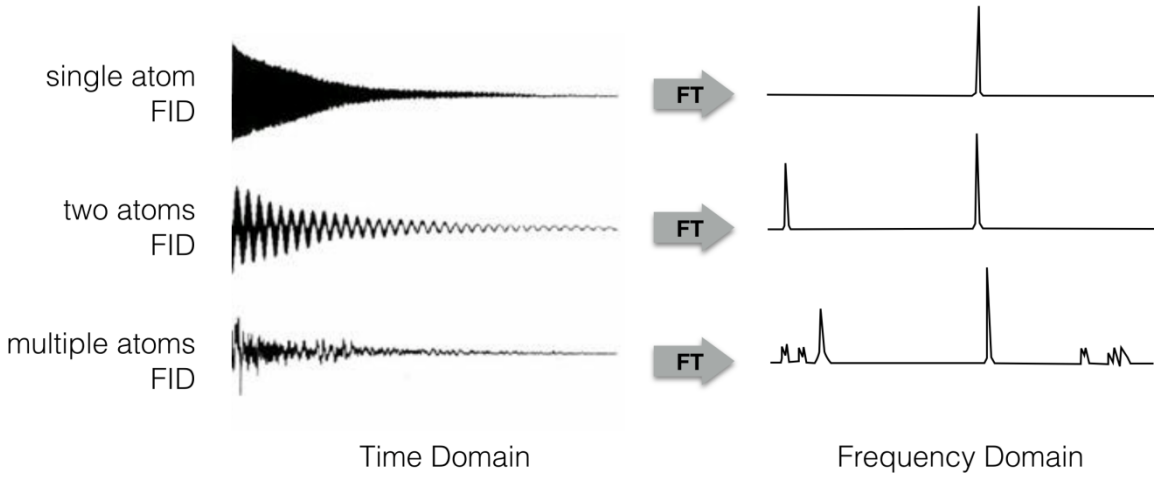

Figure 14. From the time domain to the frequency domain. Showing sample FIDs for one, two, and multiple atoms Fourier transformed to the frequency domain, the form that we read and interpret in NMR.

Details for the PqqD experiments leading to the initial deposition of peak assignments are provided in chapter 2, and final experiments leading to a PqqD solution structure and identifying binding residues for PqqD + PqqA are provided in chapter 3.

Method Selection

For our research, our method selection was guided by necessity; we chose the method that worked. The MePqqD (~10 kDa) protein initially proved so difficult to crystallize that NMR was alternatively, and successfully, employed. On the other hand, PqqC (~29 kDa), both in the form of a natural fusion (MePqqCD) and as an artificial truncation (MePqqC domain only), crystallized more readily, and their structures were determined. In general, XRC can determine larger protein structures and protein complexes, while the strength of solution NMR lies in smaller protein structures, and studies involving protein dynamics. While CEM was not employed in my dissertation research, it is emerging as the go-to method for very large and membrane-bound proteins and protein complexes.

Spin-off methods to the 'big three' methods mentioned above, such as solid state NMR and neutron diffraction crystallography, fill other specialty niches in the tool belts of structural biologists, with unique capabilities and corresponding advantages/disadvantages.

Experimental Overview

In chapters 2 and 3, solution NMR was used to determine the first physiological structure for an independent RRE (MePqqD) in a PRPS pathway, and to determine which residues of the RRE are involved in recognizing and binding the precursor peptide (MePqqA) and the first tailoring enzyme of the PQQ biosynthetic pathway, PqqE.

In chapter 4, XRC was used to determine the structure of the final enzyme in the pathway, MePqqC, from two different variant proteins: first, from the natural MePqqCD fusion, and second, from the artificial MePqqC-domain only truncation.

CHAPTER 2

This chapter is reprinted with permission.

Biomolecular NMR Assignments, 2016, ***10***(2), pp 385–389

DOI: 10.1007/s12104-016-9705-8

^{1}H, ^{13}C, and ^{15}N resonance assignments and secondary structure information for *Methylobacterium extorquens* PqqD and the complex of PqqD with PqqA

Robert L. Evans III, John A. Latham, Judith P. Klinman, Carrie M. Wilmot, Youlin Xia

Paper 1 Abstract

The ribosomally synthesized and post-translationally modified peptide (RiPP), pyrroloquinoline quinone (PQQ), is a dehydrogenase cofactor synthesized by, but not exclusively used by, certain prokaryotes. RiPPs represent a rapidly expanding and diverse class of natural products—many of which have therapeutic potential—and the biosynthetic pathways for these are gaining attention. Five gene products from the *pqq* operon (PqqA, PqqB, PqqC, PqqD, and PqqE) are essential for PQQ biosynthesis. The substrate is the peptide PqqA, which is presented to the radical SAM enzyme PqqE by the small protein PqqD. PqqA is unstructured in solution, and only binds to PqqE when in complex with PqqD. PqqD is a member of a growing family of RiPP chaperone proteins (or domains in some cases) that present their associated peptide substrates to the initial

RiPP biosynthesis enzymes. An X-ray crystal dimer structure exists for *Xanthomonas campestris* PqqD (PDB ID: 3G2B), but PqqD is now known to act as a monomer under physiological conditions. In this study, the PqqD truncation from naturally fused *Methylobacterium extorquens* (Me) PqqCD was overexpressed in *Escherichia coli* and MePqqA was chemically synthesized. Solution NMR ^{1}H-,^{15}N-HSQC chemical shift studies have identified the PqqD residues involved in binding PqqA, and ^{1}H, ^{13}C, and ^{15}N peak assignments for PqqD alone and for PqqD bound to PqqA are described herein.

Introduction (Biological context)

4,5-Dihydro-4,5-dioxo-1*H*-pyrrolo[2,3-*f*] quinolone-2,7,9-tricarboxylic acid (pyrroloquinoline quinone or PQQ), is a ribosomally synthesized and post-translationally modified peptide (RiPP) that acts as a dehydrogenase cofactor for certain alcohol and aldose sugar dehydrogenases in prokaryotes.[16, 17] Five gene products from the *pqq* operon are required for PQQ biosynthesis.[12] Two of these five products, PqqA, a 20 to 30 residue peptide with an absolutely conserved EXXXY sequence near its C-terminal end,[68-71] and PqqD, which tightly binds PqqA and forms a ternary complex with PqqA and PqqE,[44] are the focus of the NMR studies described in this paper.

PQQ is of interest for several reasons. It is a tricyclic, redox active o-quinone that is not formed by direct post-translational modification of active site residues, but instead is synthesized by way of a RiPP pathway.[15, 44, 72] PQQ is a significant antioxidant, and when present supports mitochondrial biogenesis and function in a wide range of organisms.[24-30] Additionally, PQQ demonstrates probiotic properties in mammals; studies with rats and mice have demonstrated decreased growth, reduced immune response, and declining

reproductive success when subjects were deprived of PQQ in their diets.[19-22] While initially considered a cofactor for prokaryotes only, a recent publication identified a fungal enzyme for which PQQ serves as cofactor.[18] Finally, plant studies indicate that the presence of PQQ promotes growth.[73]

Only one structural model of PqqD has been published, and this is a dimeric crystal structure of *Xanthomonas campestris* (Xc) PqqD (PDB ID: 3G2B).[63] However, the physiological state of PqqD is monomeric,[44] so the biological relevance of the XcPqqD structure is uncertain. In this study, the PqqD portion of the natural *Methylobacterium extorquens* fusion PqqCD was expressed in *Escherichia coli* to give ^{13}C and ^{15}N isotopically labeled protein, which was purified and subjected to NMR spectroscopic analysis. The interaction of the isotopically labeled MePqqD with unlabeled and chemically synthesized MePqqA, which binds with a K_d of ~ 200 nM,[44] was also probed by NMR spectroscopy. Here we present the MePqqD resonance and secondary structure assignments in the absence and presence of MePqqA in the pursuit of the physiological structure of PqqD and the mapping of the interaction surface of PqqA on PqqD.

Experimental Procedures (Methods and experiments)

Recombinant protein expression and purification

<u>*Materials*</u> - The T4 DNA ligase and restriction enzymes were obtained from New England BioLabs (Ipswich, MA). Polymerase was obtained from Agilent Technologies (Santa Clara, CA). Oligonucleotides were obtained from Eurofins (Huntsville, AL). All DNA sequencing was performed by the University of California's DNA Sequencing Facility (Berkeley, CA).

Preparation of PqqA - The unlabeled peptide, *MePqq*A ΔM1, C12S (derived from the wild type, UniProt # Q49148), was synthesized and purified to >80% purity by CPC Scientific (Emeryville, CA) and then used at that purity. The peptide sequence, KWAAPIVSEI**S**VGMEVTSYESAEIDTFN, incorporated a serine in place of the cysteine at residue position 11 to eliminate spurious dimer formation.

Preparation of ^{15}N- and ^{13}C-labeled, recombinant PqqD - The *MepqqD* gene (corresponding to amino acids 280-372 from the natural MePqqCD fusion) was cloned into the pET28a vector (EMD Millipore) using the *NdeI* and *XhoI* restriction sites. The cloned gene, incorporating an N-terminal His_6-tag, was sequence verified and used to transform *E. coli* BL21 (DE3) for gene expression. Transformed *E. coli* BL21 (DE3) cells were grown aerobically at 37 ºC in M9 minimal media supplemented with 1 g/L NH_4Cl (99% ^{15}N, Cambridge Isotopes, Tewksbury, MA), 4 g/L D-glucose (U-^{13}C, Cambridge Isotope Laboratories, Tewksbury, MA) and 50 µg/mL kanamycin. Cells were induced with 1 mM IPTG when the OD_{600} reached 0.6. Following a 12 h induction at 20 ºC, the cells were harvested by centrifugation at 6,500 rpm for 10 min. The cells were suspended in five times the mass of cell paste of 50 mM PBS (pH 7.5) and 50 mM imidazole. The cells were lysed by sonication, and the lysate was centrifuged at 20,000 rpm for 15 min. The supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare) and the column was washed at 4 ºC with lysis buffer to remove non-tagged protein, and then with 50 mM PBS (pH 6.5) and 300 mM imidazole to elute the tagged protein. The desired fractions were combined, concentrated, and buffer exchanged over PD-10 columns (GE

Healthcare) equilibrated with 25 mM phosphate buffer (pH 6.5). Yield for His_6-tagged ^{13}C-,^{15}N-labeled *MePqq*D: 27 mg/L culture.

Experimental quantities

NMR experiments were performed using D_2O matched Shigemi microtubes, 5 mm O.D. (Shigemi, Inc.). The experimental solution for PqqD alone contained 285 µL of 5.0 mg/ml (0.40 mM) ^{13}C-,^{15}N-labeled MePqqD in 25 mM potassium phosphate, pH 6.5, 7 µL of 50 mM sodium azide, and 15 µL HPLC grade D_2O (final concentrations: 4.6 mg/ml (0.37 mM) *Mex*PqqD, 1.1 mM sodium azide, 4.9% D_2O).

The experimental solution for PqqD bound to PqqA (in approximately 4-fold molar excess) was identical to the PqqD alone with the inclusion of 1.35 mg lyophilized, unlabeled MePqqA (final concentrations: 4.6 mg/ml (0.37 mM) MePqqD, 4.4 mg/ml MePqqA (1.4 mM), 1.1 mM sodium azide, 4.9% D_2O).

NMR spectroscopy

Instrumentation and conditions - All NMR data were recorded at 25°C on Bruker AVANCE™ III 850 or 900 MHz NMR spectrometers, each with 5 mm TCI CryoProbes including shielded z-gradient. Two sets of NMR data were acquired with the two samples, ^{13}C-,^{15}N-labeled PqqD and ^{13}C-,^{15}N-labeled PqqD + unlabeled PqqA. The tight binding of PqqD and PqqA precluded a titration approach. Data were processed with nmrPipe.[74] Proton chemical shifts were calibrated with respect to the water signal relative to DSS ($(CH_3)_3Si(CH_2)_3SO_3Na$); ^{15}N and ^{13}C chemical shifts were indirectly referenced to DSS.[75] Linear predictions were applied to the ^{15}N and ^{13}C dimensions to double the data size and improve digital resolution. A cosine square window function and "auto"

zero filling were applied to all ^{1}H, ^{15}N and ^{13}C dimensions. Data were analyzed with Sparky.[76]

Experiments - Sequence-specific backbone assignments were completed using AutoAssign with two 3D spectra: HNCACB and CBCA(CO)NH.[77, 78] The HNCACB creates both intra- and inter-residue correlations, whereas the CBCA(CO)NH creates only inter-residue correlations. Combining these two spectra, backbone chemical shifts, including $^{1}H_{N}$, ^{15}N, ^{13}CA, ^{13}CB, were assigned. $^{13}C'$ chemical shifts were assigned using a 3D HNCO and the first 2D ^{1}H-^{13}C plane of 3D HNCACO.[79] ^{1}H and ^{13}C side chain assignments were performed with HCCH-TOCSY, H(CCCO)NH and C(CCO)NH (mixing time: 16 ms).[80, 81] HA assignments and scalar J coupling $^{3}J_{HNHA}$ were obtained from 3D HNHA spectrum.[82] The $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ resonances of aromatic residues were assigned using 2D (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE.[83]

Results (Assignments and data deposition)

Backbone and sidechain ^{1}H, ^{15}N, and ^{13}C chemical shifts were assigned at 100% with the exception of 14 of 27 aromatic ^{13}C's (52%) and 4 of 21 aromatic ^{1}H's (19%), which were not assigned. A superposition of the 2D ^{1}H-^{15}N HSQC spectra for PqqD (blue peaks) and PqqD + PqqA (red peaks) is shown (Figure 15). From this plot, changes in the chemical shifts of ^{1}H and ^{15}N could clearly be identified. The complete backbone and side chain chemical shift assignments have been deposited in the BioMagResBank database (*www.bmrb.wisc.edu*) with accession numbers 26634 and 26690 for samples PqqD and PqqD + PqqA, respectively.

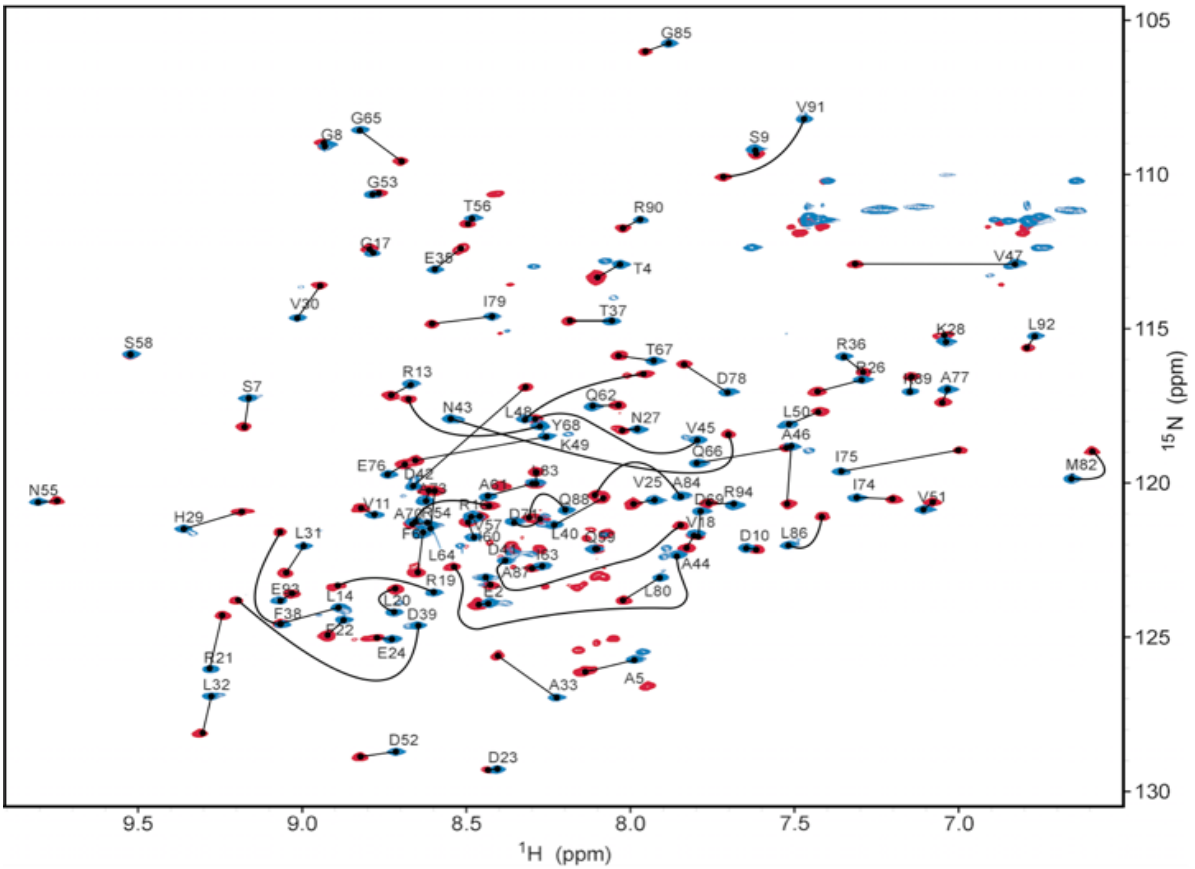

Figure 15. Overlay of 2D ^{1}H-^{15}N HSQC spectra for ^{13}C-,^{15}N-labeled MePqqD (blue) and ^{13}C-,^{15}N-labeled MePqqD + unlabeled MePqqA (red) showing peak shifts.

Discussion (Secondary structure information)

The chemical shifts of $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{13}CO resonances were used as an input to TALOS+ and CSI2.0 to predict the secondary structures of the two samples.[84, 85] The secondary structures predicted from the two methods for PqqD in each sample are very similar (Figure 16). The main difference observed between the two samples, PqqD and PqqD + PqqA, is that one β-strand from the sequence fragment RTFDL of PqqD is significantly longer in the presence of PqqA. In addition, the order parameters predicted by TALOS+ and the flexibility predicted by CSI2.0 are consistent between each sample.

Besides the N-terminus of PqqD (residues 1 and 8), the region between residues 51 and 56 in both samples indicates high mobility and disorder.

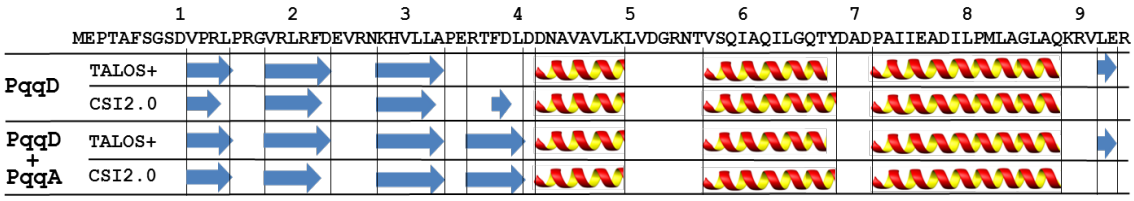

Figure 16. Secondary structures predicted by TALOS+ and CSI2.0 for MexPqqD (the first two rows) and MexPqqD + MexPqqA (the last two rows). The MexPqqD sequence is shown above the secondary structure predictions with sequence decades indicated. Arrows and helices represent β-strand and α-helices, respectively.

CHAPTER 3

This chapter (with supplemental material reintegrated) is reprinted with permission from American Chemical Society *Biochemistry*.

Biochemistry, 2017, ***56***, pp 2735–2746 DOI: 10.1021/acs.biochem.7b00247

NMR structure and binding studies of PqqD, a chaperone required in the biosynthesis of the bacterial dehydrogenase cofactor pyrroloquinoline quinone

Robert L. Evans III, John A. Latham, Youlin Xia, Judith P. Klinman, Carrie M. Wilmot

Paper 2 Abstract

Biosynthesis of the ribosomally synthesized and post-translationally modified peptide (RiPP), pyrroloquinoline quinone (PQQ), is initiated when substrate peptide, PqqA, is recognized and bound by the RiPP precursor peptide recognition element (RRE), PqqD, for presentation to the first enzyme in the pathway, PqqE. Unlike other RiPP-producing, post-ribosomal peptide synthesis (PRPS) pathways in which the RRE is a component domain of the first enzyme, PqqD is predominantly a separate scaffolding protein that forms a ternary complex with the precursor peptide and first tailoring enzyme. As PqqD is a stable, independent RRE, this makes the PQQ pathway an ideal PRPS model system for probing RRE interactions using nuclear magnetic resonance (NMR). Herein, we present both the solution NMR structure of *Methylobacterium extorquens* PqqD, and

results of $^{1}H,^{15}N$-HSQC binding experiments that identify the PqqD residues involved in binding the precursor peptide, PqqA, and the enzyme, PqqE. The reported structural model for an independent RRE, along with the mapped binding surfaces, will inform future efforts to both understand and to manipulate PRPS pathways.

Introduction

RiPPs comprise a collection of diverse natural products with probiotic as well as antibiotic, anti-cancer, and anti-viral activities.[23, 86-88] Recent genome sequencing efforts have demonstrated that RiPPs represent a major natural product biosynthesis route in addition to terpenoid, alkaloid, polyketide, and non-ribosomal peptide (NRP) biosyntheses.[11] In NRP biosynthesis, peptide bond formation requires multimodular "megaenzymes" in an assembly line encoded within the operon.[11, 34, 88] In comparison, PRPS pathways can produce equally complex and diverse molecules from simpler operons, as they utilize ribosomal machinery to produce the precursor peptide that is subsequently modified. RiPPs can encompass ≥70 amino acids, such as the putative pore-forming bacteriocin, enterocin AS-48, for which an X-ray crystal structure exists,[89] or be relatively small, as with the 330 Da tricyclic *o*-quinone, PQQ ($C_{14}H_6N_2O_8$, 4,5-dihydro-4,5-dioxo-1*H*-pyrrolo[2,3-*f*] quinolone-2,7,9-tricarboxylic acid) derived from two amino acid residues (Figure 17).[90, 91] PQQ biosynthesis is part of a PRPS biosynthesis subclass that includes pantocin, mycofactocin, and thyroid hormone biosyntheses.[11, 53, 92]

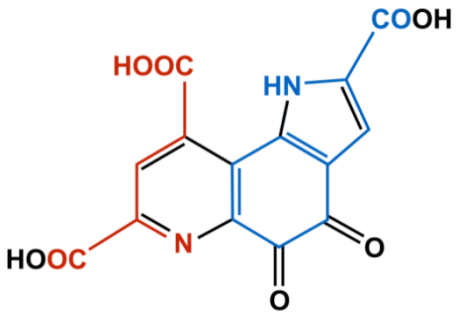

Figure 17. Chemical structure of PQQ. All of the carbons and nitrogens in PQQ come from a glutamate and tyrosine in an absolutely conserved EXXXY sequence located in the precursor peptide, PqqA. Atoms and bonds colored red are derived from the glutamate, and those colored blue are from the tyrosine. Black represents modifications during biosynthesis.

PQQ is the cofactor for a class of bacterial aldose sugar and alcohol dehydrogenases.[14-17] The cofactor is synthesized outside of the dehydrogenase framework and then non-covalently inserted into the apoenzyme.[4-6] These dehydrogenases, and therefore PQQ, give a growth advantage to the bacteria when aldose sugars and/or alcohols are available.[93] The electrons generated by the reduction of PQQ to $PQQH_2$ are fed into the constitutive electron transport chain of the organism and ultimately generate ATP. More recently, data suggested that PQQ may be a sugar oxidoreductase coenzyme in the basidiomycete mushroom, *Coprinopsis cinerea*.[18]

The majority of PQQ-producing bacteria are Gram-negative, and include a number of human pathogens, such as *Klebsiella pneumoniae* and *Burkholderia cenocepacia*.[12] Interestingly, PQQ also enhances the growth rate of some non-PQQ-producing bacteria, such as *Escherichia coli*, which under certain nutrient conditions express a PQQ-dependent dehydrogenase.[6, 94] Therefore, PQQ is a prokaryotic vitamin.

Eukaryotes do not synthesis PQQ; however, removal of PQQ from the diets of mice and rats leads to decreased fertility, smaller litter sizes, and slowed neonatal growth.[19-22] The specific mechanism(s) by which PQQ alters reproductive success is currently unknown, although data that show PQQ promotes mitochondrial metabolism are emerging.[24-30] As an antioxidant, PQQ can undergo approximately 20,000 redox cycles, making it, on a molar basis, 100 times more efficient than ascorbic acid as a cellular redox molecule.[31-33]

While biochemically intriguing in its own right, the PQQ biosynthesis pathway contains an independent RRE protein that can serve as a model system for understanding

the presentation of precursor peptides by enzyme-integrated RRE domains, which are present in > 50% of all PRPS pathways.[11, 42] PQQ biosynthesis requires five or six genes encoded by the *pqq* operon; *pqqA–E* are absolutely required, while an additional gene for a putative protease, *pqqF*, is also often present.[2, 12] PQQ biosynthesis is initiated when the precursor peptide, PqqA, is recognized and bound by PqqD for presentation to the first tailoring enzyme, PqqE, a member of the radical S-adenosyl methionine (rSAM) superfamily of enzymes that additionally contain a C-terminal subtilosin, PQQ, anaerobic sulphatase, and mycofactocin (SPASM) domain.[2] PqqD is a member of the structurally conserved RREs but is unusual in being a stand-alone peptide chaperone, as most RREs are domains within PRPS biosynthesis enzymes.[42, 95] The core segment of PqqA contains an absolutely conserved EXXXY sequence, in which the glutamate and tyrosine residues contain all the carbons and nitrogens comprising PQQ (Figure 17).[96, 97]

This paper presents the nuclear magnetic resonance (NMR) solution structure for *Methylobacterium extorquens* (Me) PqqD, which represents the first experimental structure of PqqD in its physiological monomeric form. Further, we identify the residues of PqqD involved in binding PqqA alone, and in binding PqqE within the ternary complex. These findings are placed in the context of other structural work on RRE domain interactions with RiPP precursor peptides and the enzymes of which they are a part.

Experimental Procedures

Materials

The T4 DNA ligase and all restriction enzymes were purchased from New England BioLabs (Ipswich, MA). Oligonucleotides were purchased from Eurofins (Huntsville, AL). Polymerase was obtained from Agilent (Santa Clara, CA). DNA sequencing was performed by the University of California DNA Sequencing Facility (Berkeley, CA). The precursor peptide, MePqqA (Uniprot entry Q49148), was chemically synthesized and purified (80% purity) by CPC Scientific (Emeryville, CA) and used without further purification in experiments. The MePqqA peptide used in these studies did not include the N-terminal methionine, and residue 11 was changed from a cysteine to a serine to avoid interpeptide crosslinking. The final sequence for the construct was *KWAAPIVSEISVGM**EVTSY**ESAEIDTFN* (the absolutely conserved EXXXY sequence is highlighted).

Recombinant expression of ^{13}C, ^{15}N-labeled PqqD

Although PqqD predominantly exists as a separate protein in most PQQ biosynthesis pathways, MePqqD is found N-terminally fused to MePqqC. A truncation of MePqqCD (Uniprot entry Q49150) containing only the MePqqD portion was constructed, which had been done in previous studies in which the presence of the MePqqC portion of the native polypeptide was shown to have no effect on the binding of MePqqA and MePqqE to MePqqD.[44] The truncation sequence was derived by aligning the MePqqCD natural fusion sequence with PqqD and PqqDE sequences from 8 other bacterial species: *Pseudomonas putida*, *Acinetobacter calcoaceticus*, *Klebsiella pneumoniae*, *Azotobacter*

vinelandii, *Gluconobacter oxydans*, *Rhodopseudomonas palustris*, *Methylocystis sp.* (containing the natural fusion, PqqDE), and *Xanthomonas campestris* (XcPqqD)(Clustal Omega, www.ebi.ac.uk/Tools/msa/clustalo/).[98-100] The 94 residues comprising the MePqqD used in the study include an N-terminal methionine followed by 5 residues (*EPTAF*) representing the juncture of the linker and PqqD protein and the remaining C-terminal residues from MePqqCD. Thus, the MePqqD sequence is: *MEPTAFSGSDVPRLPRGVRLRFDEVRNKHVLLAPERTFDLDDNAVAVLKLVDGRNTV SQIAQILGQTYDADPAIIEADILPMLAGLAQKRVLER* (Figure 18).

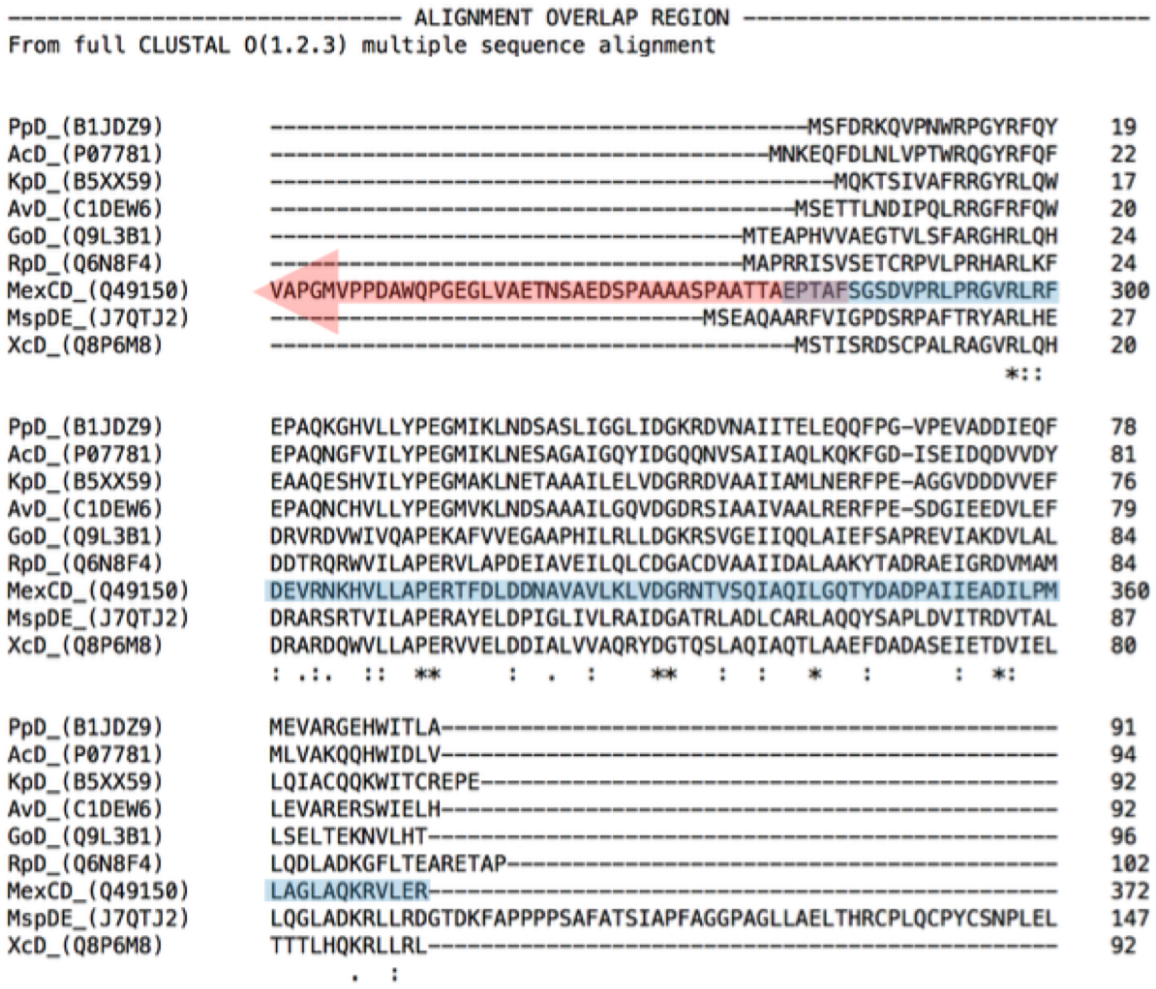

Figure 18. The PqqD portion of MePqqCD was identified using an alignment of 9 species, 2 of which were natural fusions. 'MexCD_(Q49150)' represents MePqqCD. The pink/purple arrow identifies the linker region and PqqC enzyme portion of the fusion, and the purple/blue, PqqD. The Uniprot identifiers are enclosed in parentheses and the alignment was completed using Clustal Omega (http://www.uniprot.org/ and http://www.ebi.ac.uk/Tools/msa/clustalo/, respectively).

The truncated Me*pqqD* gene construct was cloned into a kanamycin resistant pET28a vector with an N-terminal His_6-tag (EMD Millipore). The gene was used to transform *E. coli* BL21 (DE3) cells for gene expression. Cells were grown aerobically in minimal media supplemented with ^{13}C- and ^{15}N-labeled ammonium chloride and D-glucose,

respectively (Cambridge Isotope Laboratories, Tewksbury, MA). The yield for the His_6-tagged, ^{13}C-,^{15}N-labeled MePqqD was 27 mgs per liter of culture. After immobilized metal affinity chromatography purification (nickel column), the protein was buffer exchanged into 25 mM potassium phosphate (pH 6.5) for all NMR experiments. For experiments involving the ternary complex only, the PqqD sample had been lyophilized for storage.

Recombinant expression of unlabeled PqqE

The gene encoding MePqqE (UniProt entry P71517) was cloned into the pET28a vector from genomic DNA using the NdeI and XhoI restriction sites. After sequence verification, the cloned gene was used to transform *E. coli* BL21(DE3) that also contained a *suf* operon plasmid, pPH149, for gene expression. The pPH149 and *pqqE*/pET28a-transformed cells were grown aerobically at 37 °C in TB media containing 35 g/ml chloramphenicol and 50 g/ml kanamycin. At a cell density of A_{600} = 0.6, 50 μM iron(III) citrate and 5 mM fumarate were added. The growth flasks were immediately stoppered so that the transition to anaerobic conditions occurred. Anaerobic growth was allowed to continue for 30 minutes at 37 °C, after which the expression of His_6-MePqqE was induced by adding 0.4 mM isopropyl-D-galactopyranoside. Following a 12-hour induction period at 19 °C, cells were centrifuged for 10 minutes at 6,500 rpm and then frozen. Subsequent work was performed in an anaerobic chamber. Cells were suspended in 5x (mass of cell paste) degassed 50 mM Tris buffer (pH 7.9), 50 mM imidazole, 1 mM TCEP, and 200 mM sodium chloride (lysis buffer). Cells were lysed following the manufacturer's protocol for BugBuster (Novagen). BugBuster was supplemented with

Benzonase (Novagen). Lysate was transferred into sealed tubes for centrifugation (outside the anaerobic chamber) for 15 minutes at 20,000 rpm. The sealed tubes were then moved back into the anaerobic chamber where the combined supernatants were loaded onto a 5 ml His-Trap FF column (GE Healthcare) using a peristaltic pump. The column, maintained at 22 ºC, was washed first with lysis buffer to remove non-tagged protein and then with a degassed solution consisting of 300 mM imidazole, 50 mM Tris (pH 7.9), 1 mM TCEP, and 200 mM sodium chloride to elute the tagged protein. Fractions were combined, concentrated, and buffer exchanged using PD-10 columns equilibrated with degassed 50 mM Tris (pH 7.9), 100 mM sodium chloride, 5% glycerol, and 1 mM TCEP. The protein, anaerobically aliquoted and sealed in cryogenic vials, was then flash frozen in liquid nitrogen. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was used to confirm protein homogeneity. The yield of His_6-MePqqE was 18 mgs per liter of culture.

NMR Experiments

The PqqD protein was highly stable, and periodic $^{1}H,^{15}N$-HSQCs on the single sample used for all experiments showed no discernable degradation over the multiple months of data collection. Likewise, the varying PqqD:PqqA and PqqD:PqqA:PqqE ratios were highly stable, and a single preparation of each of these also lasted for the duration of the experiments needed for structure determination and binding studies. Protein samples were stored at 4 ºC when not in use.

All NMR experiments were performed using D_2O matched 5 mm Shigemi microtubes. The PqqD sample contained 0.14 mM MePqqD, 1.0 mM sodium azide, and

5% D_2O. The physiological stoichiometry for PqqD with PqqA is 1:1, with a dissociation constant (K_D) of ~200 nM for the MePqqA-PqqD complex.[44] To determine the proper molar ratio of PqqD and PqqA for defining the residues of PqqD affected by binding PqqA, three two dimensional (2D) $^{1}H,^{15}N$-HSQC spectra were acquired at PqqD:PqqA ratios of 1:0, 1:0.7, and 1:1.3. The 2D HSQC spectra demonstrated that at a PqqD:PqqA ratio of 1:1.3, all chemical shift changes were complete with peaks fully evolved. The PqqD–PqqA samples were made by adding 0, 5, and 10 μL of 6.4 mM PqqA, in identical buffer, to the 270 μL PqqD only sample (for the 1:0, 1:0.7, and 1:1.3 sample ratios, respectively). The starting concentration of PqqD was 0.17 mM in the presence of 1.0 mM sodium azide, and 5% D_2O.

The physiological stoichiometry for PqqA with PqqE and PqqD is 1:1:1, with K_D values of ~12 μM for the MePqqD–PqqE binary complex and ~5 μM for the MePqqD–PqqA–PqqE ternary complex.[44] Protein solubility was an issue with the PqqD–PqqA–PqqE sample, as PqqE tended to precipitate out of solution as the ratio of PqqE to PqqD exceeded 0.2 to 1. However, chemical shift assignments for the identification of PqqD residues involved in binding PqqE in the presence of PqqA could be ascertained using a 1:1.25:0.2 PqqD:PqqA:PqqE ratio. Therefore, the volume of the PqqD–PqqA–PqqE sample was made to be similar to those of the other samples but contained 0.80 mM MePqqD (~ six times the quantity used in the previous experiments in order to increase signal-to-noise due to the weak NMR signal of the slower tumbling ternary complex), 1 mM MePqqA (25% excess over PqqD), 0.16 mM MePqqE (only 20% of PqqD), 1.0 mM sodium azide, and 5% D_2O. The PqqE was diamagnetic as the samples were treated

aerobically throughout all experiments, as was done in the previously published work that determined K_D values.[44]

All experiments were performed at 25 °C on Bruker 850 or 900 MHz spectrometers, each equipped with 5 mm TCI CryoProbes. NMRPipe[74] was used to process experimental data. Proton chemical shifts were calibrated with respect to the water signal relative to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), and ^{13}C and ^{15}N chemical shifts were indirectly referenced to DSS.[101] Data were analyzed leading to peak assignments using Sparky.[76] The complete backbone and side chain chemical shift assignments for MePqqD have been deposited in the BioMagResBank (BMRB)(*www.bmrb.wisc.edu*) with accession numbers 26634 for PqqD only, 26690 for the PqqD–PqqA complex, and 26696 for the PqqD–PqqA–PqqE ternary complex.[102, 103]

Binding studies (PqqD and PqqA)

As for chemical shift assignments, NMR data acquired from PqqD and PqqA were similar to those of PqqD alone. Typically, sequence-specific backbone assignments were completed using two three dimensional (3D) spectra, IBS_Best_HNCACB and IBS_Best_HN(CO)CACB.[104] The IBS_Best_HNCACB experiment reveals both intra- and inter-residue correlations, whereas the IBS_Best_HN(CO)CACB experiment reveals inter-residue correlations. The IBS_Best version of the two experiments improves signal sensitivity. Via combination of these two spectra, backbone chemical shifts, including $^{1}H_N$, ^{15}N, ^{13}CA, and ^{13}CB were assigned.[102] The assigned $^{1}H_N$ and ^{15}N shifts were then used to identify PqqD residues involved in the binding of PqqA. Data sets acquired from experiments using both PqqD alone and PqqD–PqqA samples were used to assign $^{13}C'$

and sidechain chemical shifts.[102] Complete chemical shift assignments were obtained for both samples. Finally, an overlay of the $^{1}H,^{15}N$-HSQC spectra for PqqD alone and PqqD with PqqA identified residues for which the weighted peak shifts (Equation 1)[105] were at least 1.5 standard deviations from the statistical mean for all shifts, thus indicating which of the RRE residues are involved in binding the precursor peptide.

$$\Delta\delta_{weighted} = \sqrt{\left(\frac{\Delta\delta_{^{1}H}}{1}\right)^2 + \left(\frac{\Delta\delta_{^{15}N}}{5}\right)^2} \quad \textbf{(1)}$$

Binding studies (PqqD–PqqA–PqqE)

A comparison of the 2D $^{1}H,^{15}N$-HSQC spectra acquired for PqqD, PqqD–PqqA, and PqqD–PqqA–PqqE samples showed that the changes in chemical shifts for spectra from the binary complex to the ternary complex were not as significant as those from PqqD alone and PqqD with PqqA. It was not, therefore, necessary to acquire a full set of NMR data to assign the chemical shifts of PqqD in the ternary complex as it was for PqqD with PqqA. We were able to use the assigned chemical shifts of PqqD in the binary complex and limited NMR data for the PqqD–PqqA–PqqE sample to assign the backbone chemical shifts of PqqD in the ternary complex. These data included a 3D IBS_Best_HNCA, a 3D HNCO, and the first ^{1}H-^{13}C 2D plane of an HN(CA)CO, allowing for the assignment of chemical shifts for backbone H_N, N, CA, and CO atoms.[104] No experiments that aimed to assign sidechain chemical shifts were performed. A 2D $^{1}H,^{15}N$-HSQC spectrum using the IBS_SOFAST pulse sequence was also acquired.[106] The IBS_SOFAST pulse sequence enhanced signal sensitivity to about three times that of a regular HSQC and allowed for the acquisition of a 2D HSQC spectrum with high sensitivity in just 20 minutes. Again, an overlay of the $^{1}H,^{15}N$-HSQC spectra

for PqqD alone versus that of the PqqD–PqqA–PqqE sample identified residues for which the weighted peak shifts were at least 1.5 standard deviations from the mean for all shifts, indicating RRE residues involved in binding the initial tailoring enzyme, PqqE, during peptide presentation.

3D structure determination

To determine the 3D structure of MePqqD, we utilized internuclear distances from NOE spectra, dihedral angles estimated from the chemical shifts using TALOS+, scalar $^{3}J_{HNHA}$ coupling constants from HNHA spectrum, and hydrogen bonds (Table 3).[84, 107] Even though the only deposited structure for PqqD [XcPqqD, Protein Data Bank (PDB) entry 3G2B] is a non-physiological dimer in which β-hairpins are exchanged between monomers, an *in silico* PqqD monomer model was created using PyMOL by combining the β-hairpin from one monomer with the α-helical bundle of the other (Figure 19).[63]

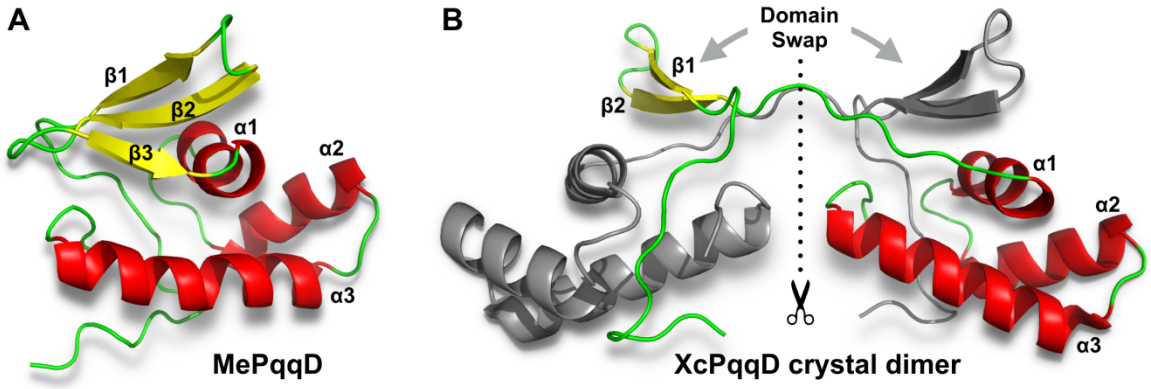

Figure 19. The PqqD structure. (A) The MePqqD monomeric NMR solution structure. (B) The XcPqqD dimeric crystal structure showing that the domain swapped β-hairpins occupy a similar position in the MePqqD monomer.

This model had a fold similar to the RRE domains of PRPS biosynthesis enzyme crystal structures and was used to inform NOE peak assignments. The intensities of NOE cross

peaks of two 3D ^{15}N- and ^{13}C-edited NOESY spectra were binned into three categories: strong, medium, and weak, corresponding to 1.8–3.7, 1.8–5.0, and 1.8–6.0 Å distance restraints, respectively. A total of 2,697 NOE-derived distances were obtained from the NMR spectra and converted into unambiguous structural restraints. On average, we obtained 28.7 unique distance restraints per residue. The chemical shifts for the $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}CO resonances were used as an input for TALOS+ to predict backbone dihedral angles ϕ and ψ. A total of 158 ϕ and ψ angle restraints were implemented in the structure calculation protocol. Three-bond scalar $^{3}J_{HNHA}$ coupling constants were obtained from the 3D HNHA experiment using the equation $J = [(-S_{cross}/S_{diagonal})/(2\pi\delta)]^{1/2}$, where S_{cross} and $S_{diagonal}$ are the signal intensities of cross and diagonal peaks, respectively, and δ (= 0.013 s) is the time delay used in this experiment.[107] A total of 82 $^{3}J_{HNHA}$ coupling constants were entered into the structure calculation protocol.

Structure calculations were performed using XPLOR-NIH version 2.37.[108, 109] As a starting structure, an extended conformation of the polypeptide with randomized dihedral angles was used. Torsion angle restrained molecular dynamics (rMD) was performed at 3,500 K for 100 ps. Subsequently, hybrid simulated annealing and rMD calculations were performed. After several rounds of structure calculations, the conformers converged to a similar fold with an ~1 Å root mean square deviation (RMSD) for all heavy atoms. At this stage, 15 hydrogen bonds between the carbonyl and the amide groups of the backbone in the α-helices and between β-strands were identified using two criteria: donor-acceptor distances of < 2.4 Å and angles of < 35°. Therefore, a total of 30

hydrogen bond restraints (each hydrogen bond having two distance restraints) were included in the final stage of energy minimization. One hundred conformers were calculated, and the 20 best structures were chosen for the final analysis. During structural refinement, calculated structures and constraints were validated using the Protein Structure Validation Suite (PSVS) (http://psvs-1_5-dev.nesg.org/). The average RMSDs for all atoms and for just the α-carbons were obtained using the PyMol "super" command (allowing 5 outlier rejection cycles).[110] The calculations were performed using the lowest energy MePqqD model's core fold (residues 17–90) and the XcPqqD *in silico* model's core fold (defined as residues 15–31 from one monomer and residues 34–88 from the second)(Figure 19). The NMR solution structure for MePqqD, based on BMRB entry 26634, has been deposited as PDB entry 5SXY.[111, 112] Figures containing images of protein structures were created using MacPyMOL,[113] and NMR HSQC spectra were generated using Sparky.[76]

Table 3. Structural statistics for 20 model ensemble

NMR Restraints	2,967
Total distance restraints Intraresidue (i-j=0) sequential (|i-j|=1) medium range (1<|i-j|<5) long range (|i-j|5)	2,697 549 862 740 546
Total dihedral angle restraints φ ψ	158 79 79
$^{3}J_{HNHA}$ coupling constants	82
Hydrogen bonds	30
RMSD from idealized geometry bonds (Å) angles (deg.)	 0.012 1.4
RMSD (Å) all backbone atoms of residues 7-93 all heavy atoms of residues 7-93 all backbone atoms of ordered residues* all heavy atoms of ordered residues	 0.7 ± 0.1 1.3 ± 0.2 0.6 ± 0.1 1.1 ± 0.2
Ramachandran Most favorable regions Allowed regions Disallowed regions	 97.9 % 2.1 % 0 %

*Ordered residues: 7-21, 29-32, 39-61, 71-86, and 91-93.

Results

Structure of M. extorquens PqqD

This structure represents the first detailed physiological structure of PqqD in its monomeric solution state.[44, 63, 114] The restraints used in the structure determination and a comparison of the 20 lowest-energy structures are given in Table 2. Labeled PqqD showed an excellent ^{1}H,^{15}N-HSQC signal-to-noise ratio with clear peak separation even in the most peak-dense regions of the spectra (Figure 20). All backbone and sidechain ^{1}H, ^{15}N, and ^{13}C chemical shifts were assigned with the exception of 14 (of 27) aromatic ^{13}C

atoms and 4 (of 21) aromatic ^{1}H atoms, because of the inherent ambiguity of C-H definition within an aromatic ring.[102] On the basis of $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}CO chemical shift resonances, PqqD was predicted to contain three β-strands and three α-helices, and the top 20 (lowest-energy) annealed structures display a common core containing precisely these elements (Figure 21).

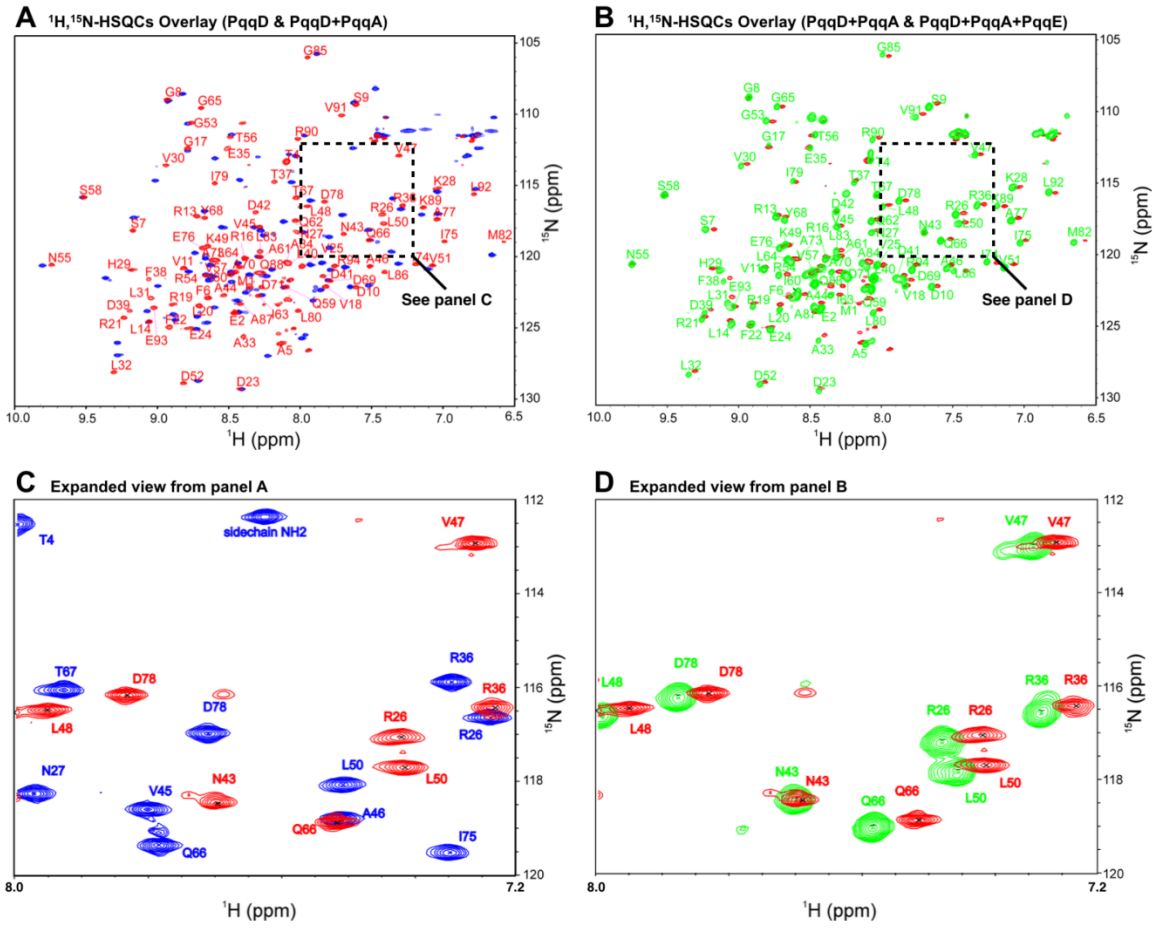

Figure 20. (A) $^{1}H,^{15}N$-HSQC of PqqD alone (blue peaks) overlaid with the $^{1}H,^{15}N$-HSQC of PqqD in complex with PqqA (red peaks). (B) $^{1}H,^{15}N$-HSQC of PqqD in the binary complex with PqqA (red peaks) overlaid with that of PqqD in the ternary complex with PqqA + PqqE (green peaks). (C) Expanded section from panel A showing peak separation. (D) Expanded section from panel B showing peak separation.

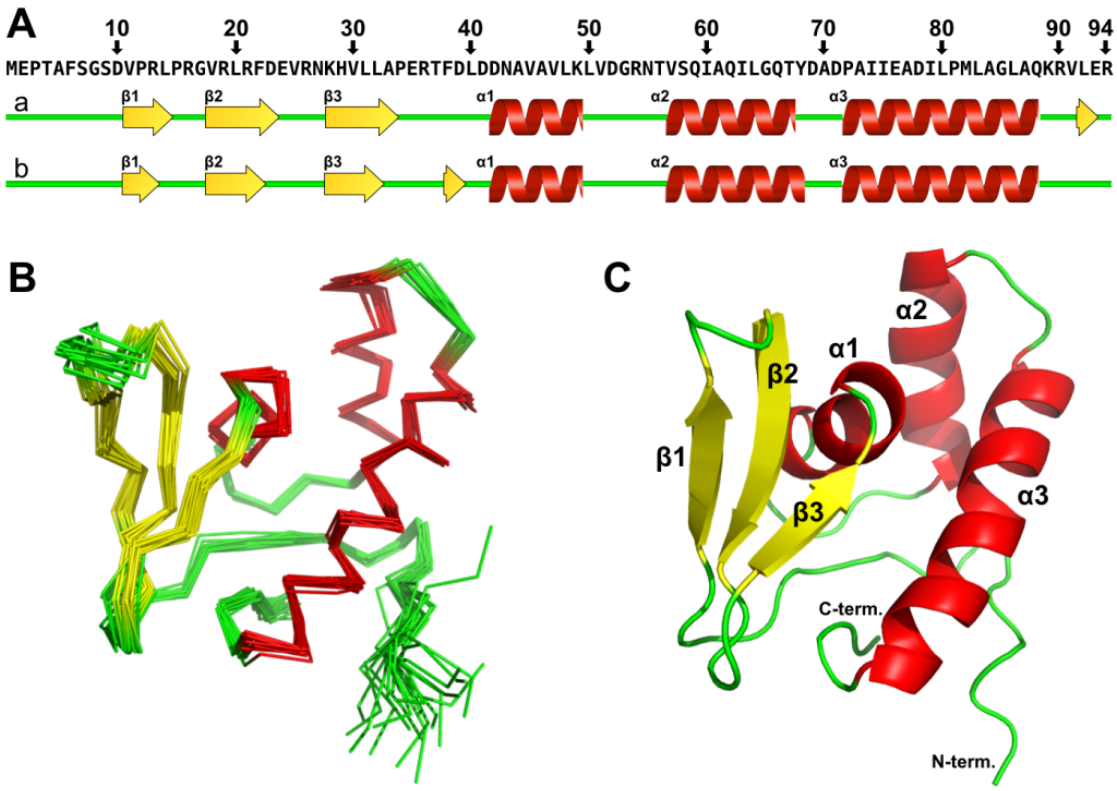

Figure 21. NMR structure of MePqqD. (A) Topology and location of secondary structure vs primary sequence as predicted by TALOS+ (a) and CSI2.0 (b). (B) Cα trace of the superposition of the 20 lowest energy NMR conformers. (C) Cartoon of the lowest energy conformer. α-helices, red; β-strands, yellow; loops, green.

All residues in the ensemble (20 x 94) were in favored or allowed Ramachandran regions with no residues in disallowed regions. The PqqD structural order predicted by TALOS+[84] and CSI2.0[85, 115] indicated a high level of disorder at the N-terminus [residues

1–7, (Figure 22)].

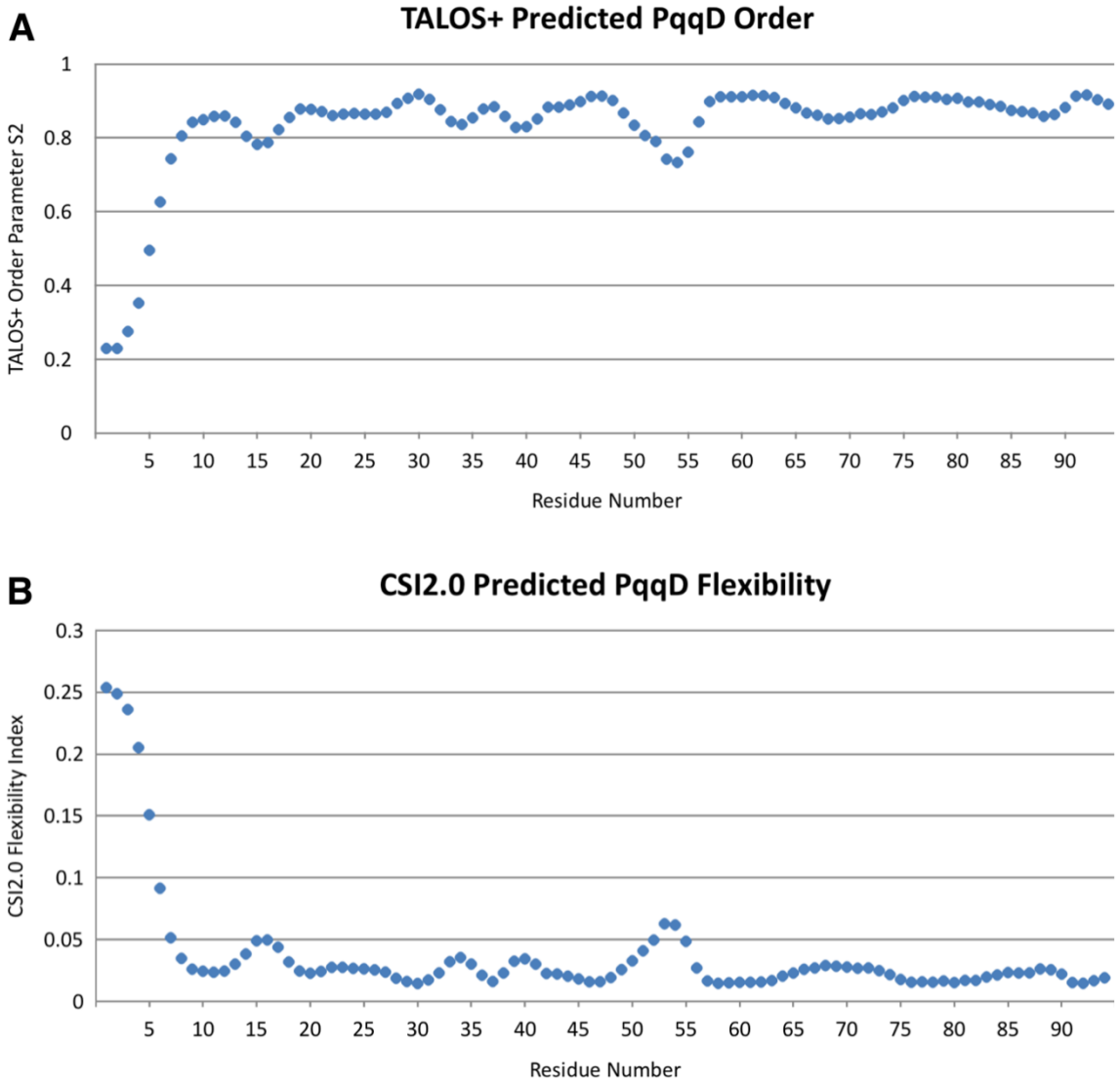

Figure 22. Structural order and flexibility by residue number as predicted by TALOS+ (A) and CSI2.0 (B).

This is perhaps not surprising as the PqqD construct used in this study was derived from the natural MePqqCD fusion, so the N-terminus in the native protein is connected to the

C-terminus of PqqC via a 26-residue polypeptide linker (Figure 23).

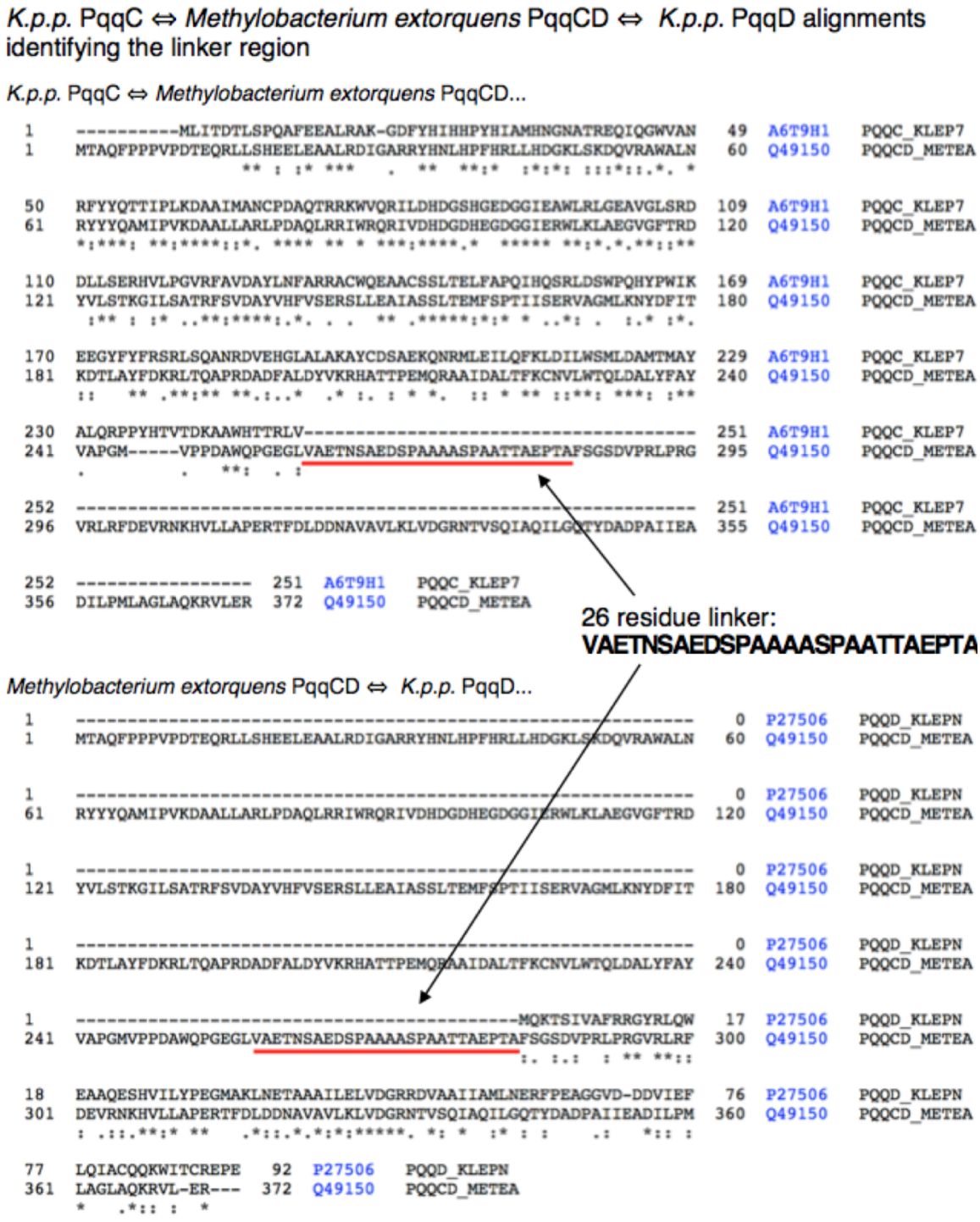

Figure 23. The linker region in MePqqCD was identified using an alignment of MePqqCD and *Klebsiella pneumoniae* PqqC and D.

There is no evidence that PqqC and PqqD specifically interact, so MePqqD can be viewed as "tethered" and independent of MePqqC.

PqqD residues involved in binding PqqA

PqqA must be bound to PqqD to become a substrate for the first enzyme in PQQ biosynthesis, PqqE.[45] To map the binding site, NMR was used to define the PqqD residues with which PqqA interacts. 2D HSQC spectra demonstrated that at a PqqD:PqqA ratio of 1:0.7 (Figure 24B), the chemical shifts observed were in positions identical to those for PqqD alone (Figure 24A) and PqqD fully bound to PqqA (Figure 24C). This indicates slow chemical exchange relative to the NMR time scale and is consistent with the reported K_D of ~200 nM.[44] As such, an HSQC titration experiment to track migrating peaks was not possible, and all chemical shifts for the PqqD–PqqA sample had to be completely reassigned.

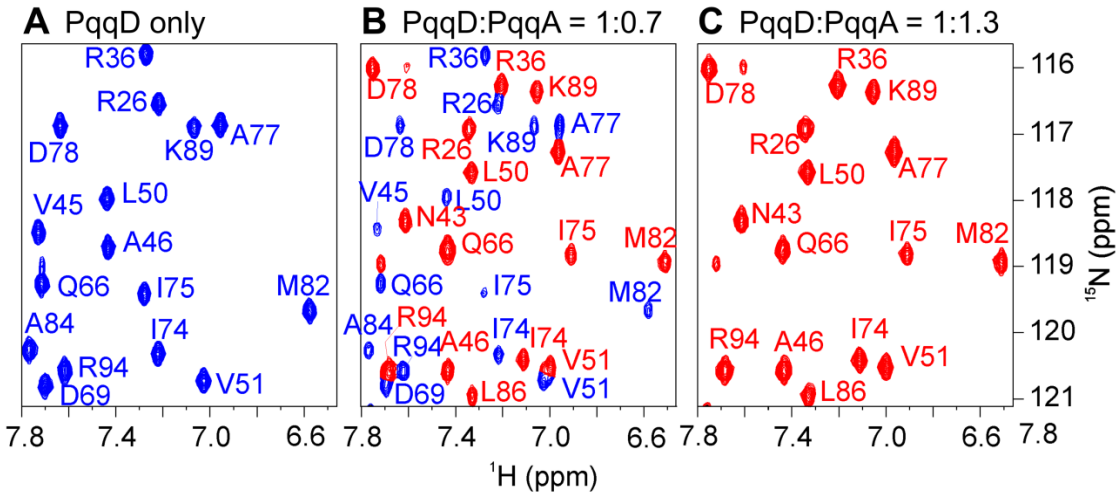

Figure 24. (A) .^{1}H,^{15}N-HSQC for PqqD alone. (B) An overlay of the spectra before saturation at a PqqD:PqqA ratio of 1:0.7. (C) The fully evolved spectrum indicating saturation at a ratio of 1:1.3.

An examination of the weighted (Equation 1) $^{1}H,^{15}N$-HSQC peak shifts ≥ 1.5σ above the mean shift between the unbound and binary complex spectra identified the following PqqD residues as being involved (either directly or indirectly) in binding PqqA: Phe38*, Asp39*, Asp41*, Asp42*, Asn43*, Ala44*, Val45, Val47, Leu48, and Val91 (asterisks indicate residues ≥ 2σ above the mean shift). These residues are almost contiguous in the primary sequence (38–48), except for residue 91, and those ≥ 2σ above the mean shift are located in the third β-strand (β3) and first α-helix (α1) (Figure 21C) that together form a "saddle" in the structure between the β-sheet and the α-helical bundle (purple, Fig. 25).

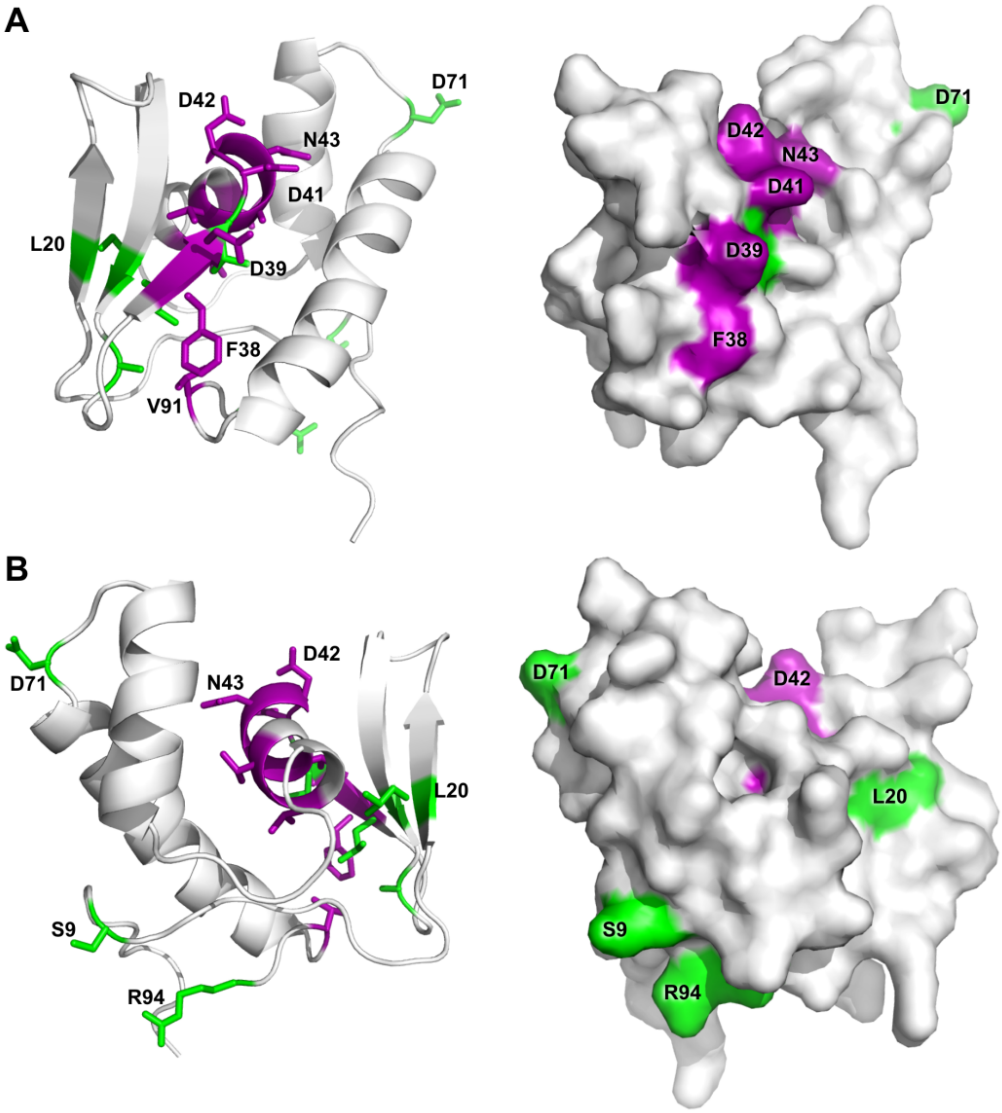

Figure 25. PqqD residues identified as having $^{1}H,^{15}N$-HSQC peak shifts ≥ 1.5σ above the mean shift are indicated for the complex with PqqA (purple) along with additional residue peak shifts upon addition of PqqE to the PqqD + PqqA complex (green). Ternary complex interactions are based on chemical shifts of the backbone only, however sidechains are displayed for recognition ease. PqqD cartoon and molecular surface renderings from the same viewpoint are shown in each panel, with panel B being a 120° rotation of panel A around an axis vertical in the plane of the paper.

Additional PqqD residues perturbed when PqqE binds the PqqD–PqqA binary complex

For the rSAM enzyme, PqqE, to act on the precursor peptide, PqqA, it must be bound to PqqD.[45] PqqE is the first enzyme in PQQ biosynthesis, and it catalyzes the formation of a carbon–carbon cross-link between the Tyr and Glu of PqqA, the residues that provide most of the raw material for PQQ (Figure 17).[45, 114] Although MePqqE can form a binary complex with MePqqD in the absence of MePqqA (K_D ~ 12 μM), it is the ternary complex (PqqD–PqqA–PqqE) that is physiologically relevant.[44, 45, 114] Therefore, NMR was used to define additional residues on PqqD that were involved in binding PqqE within the context of the ternary complex. PqqE is large compared to PqqD and PqqA (MePqqE, 41.7 kDa; MePqqD, 10.4 kDa; MePqqA, 3.07 kDa), so the ternary complex tumbles significantly more slowly than PqqD–PqqA. To overcome the resulting weaker signal-to-noise ratio, the concentration of PqqD was increased, but this led to PqqE precipitating out of solution at ratios greater than 20% PqqD. Therefore, while binding of PqqE to the PqqD–PqqA complex was within the regime where an NMR titration experiment could be possible (K_D ~5 μM), it was impractical.[44] Even though PqqD, PqqA, and PqqE bind in a 1:1:1 ratio, PqqE at 20% of the concentration of PqqD was sufficient to visualize HSQC peak shifts (Figures 20B, D). Reassignment of the chemical shifts was performed for backbone atoms only.

$^{1}H,^{15}N$-HSQC peak shifts ≥ 1.5σ above the mean shift between the binary complex and ternary complex spectra indicated that the additional PqqD residues involved in binding PqqE in the presence of PqqA are: Ser9*, Leu20, Leu31*, Ala33*, Leu40, Asp71, and Arg94* (asterisks indicate residues ≥ 2σ above the mean shift) (green, Fig.

25). The two most N-terminal β-strands and residues in the N-terminal and C-terminal tails of the polypeptide are interacting (directly or indirectly) with PqqE. Interestingly, Leu40, which showed no peak shift in the PqqD–PqqA complex but lies between the β-strand and α-helix to which PqqA binds, now becomes perturbed in the ternary complex (green, Fig. 25). Whether this is a direct interaction, or an indirect effect mediated through PqqA, is currently unknown.

Discussion

The monomeric NMR structure of MePqqD represents the first detailed, physiologically relevant structure of PqqD

Size exclusion chromatography, along with small angle x-ray scattering (SAXS) experiments, and now NMR show that PqqD is physiologically monomeric.[44, 114] The MePqqD NMR solution structure closely resembles the *in silico* models created to model the SAXS data and to show that PqqD likely belongs to the "prevalent peptide-binding fold" in PRPS biosynthesis pathways.[42, 44] These *in silico* models, and the model used to assist with NOE peak assignments, were informed by an earlier x-ray structure (PDB entry 3G2B) showing the protein as a dimer.[63] The XcPqqD dimerization observed in the structure was likely due to the high protein concentration required for crystallization.[63] While the secondary structure is largely intact in the XcPqqD crystal structure, the first two β-strands from the β-sheet subdomain of one monomer have switched with the equivalent subdomain portion of the second monomer. Interestingly, the subdomain-swapped elements from the XcPqqD crystal dimer retain very similar packing between the α-helices and β-sheet as observed in the MePqqD monomer (Figure 19). The RMSD

for α-carbons of the XcPqqD core fold (β1–β2 from one monomer and β3, α1–α3 from the second) and the MePqqD core fold was 1.72 Å. This indicates remarkable similarities of the fold between the NMR structure for MePqqD and the subdomain-swapped monomer from the XcPqqD crystal structure.

The β-strands and α-helices in MePqqD also closely match the constituent RRE folds seen in three PRPS biosynthesis enzymes for which crystal structures in complex with peptide exist: *Lactococcus lactis* NisB (PDB entry 4WD9), a lanthipeptide dehydratase in the biosynthesis pathway of nisin;[116] *Lyngbya sp. PCC-8106* LynD (PDB entry 4V1T), a fused cyclohydratase involved in cyanobactin biosynthesis;[117, 118] and *E. coli* MccB (PDB entry 3H9Q), an adenylating enzyme in the microcin C7 pathway[119] (Figure 26).

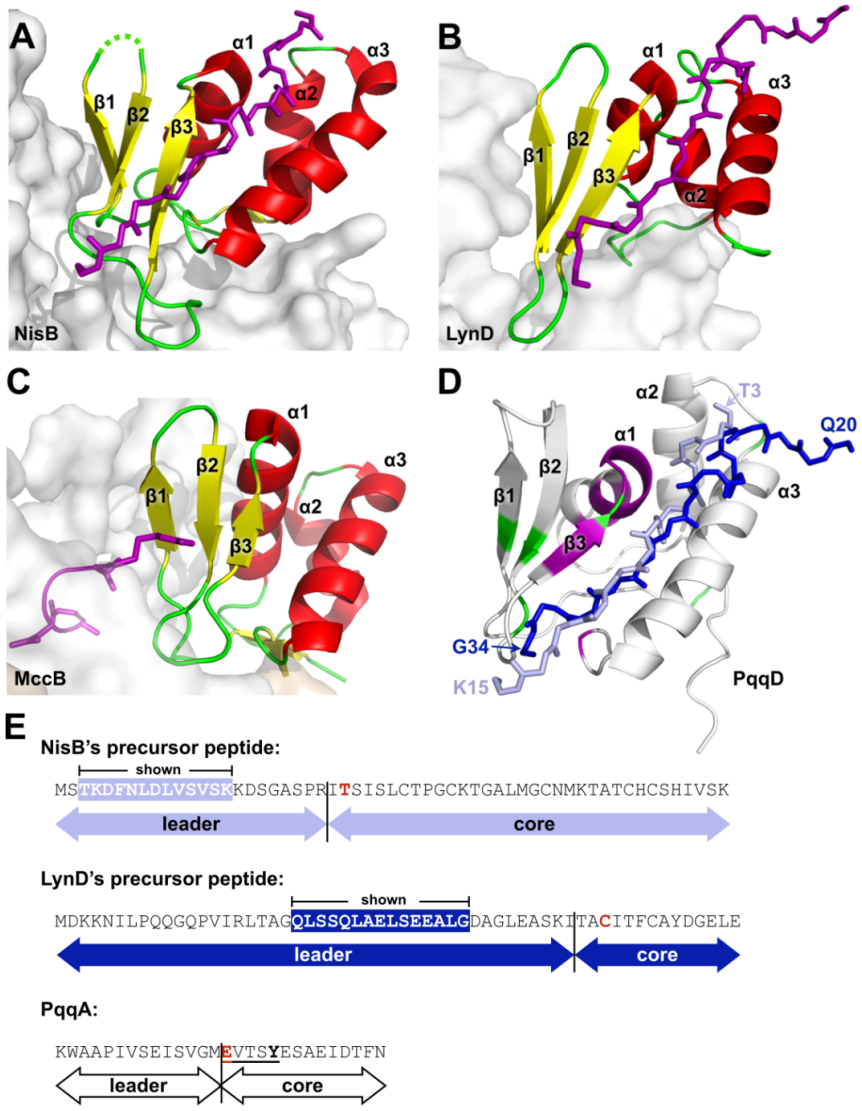

Figure 26. Comparison of PqqD to RREs from other PRPS enzymes in complex with peptide. (A) NisB, PDB 4WD9 (residues 142–223), (B) LynD, PDB 4V1T (residues 1–81), and (C) MccB, PDB 3H9J (residues 1–78). Constituent RREs with bound peptides are shown as cartoon and colored by secondary structure with bound precursor peptides drawn as stick and colored purple. The enzyme of which the RRE is a part is shown as a gray molecular surface. (D) PqqD residues perturbed by binding PqqA and the rSAM enzyme, PqqE, are colored purple and green, respectively. The peptides of NisB and LynD crystal structures are overlaid based on superposition of the RREs with PqqD (NisB peptide, light blue; LynD peptide, dark blue). (E) Precursor peptide sequence comparison between NisB, LynD, and PqqD. The peptide residues observed in the NisB and LynD crystal structures are white in colored boxes. The most proximal residue that is post-translationally modified is colored red.

PqqA binding to PqqD

Chemical shift perturbations observed during substrate binding can be caused directly by noncovalent interactions with either solvent molecules or binding partners, or indirectly by alterations to the conformational environment of residues not in direct contact. Via comparison of the residues that showed the greatest chemical shifts in the PqqD–PqqA $^{1}H,^{15}N$-HSQC spectrum with the peptide positions in the NisB, LynD, and MccB crystal structures, PqqA appears to bind to a region on the surface of PqqD similar to that of the peptides of NisB and LynD (Figure 26D).[116, 120] In the case of MccB, the substrate, MccA, is a heptapeptide, one of the shortest RiPP substrates known, and the short length does not allow substantial interactions with the MccB RRE (Figure 26C).[119] As such, the MccB–MccA interaction is likely atypical with contact limited to one end of the RRE β-sheet, principally through the side chain of MccA Arg2. In the cases of NisB and LynD, the peptide forms an additional β-strand to the core β-sheet with the expected main chain hydrogen bonds to β3, and this constitutes the defining, major interaction with the RRE. In the case of PqqD, the β3 strand chemical shifts are significantly perturbed, suggesting that PqqA may also form a β-strand with the PqqD β-sheet. The residues prior to those within the β-strand adopt different conformations in the NisB and LynD peptides (Figure 26). Despite this, both the NisB and LynD peptide side chains in this region interact with the N-terminal α3 helix side chains and primarily involve bulky hydrophobic residues, such as Leu, Tyr, and Phe. However, it should be noted that the first four ordered residues of the LynD peptide are in a substantial crystal contact that may influence the conformer observed in this particular case. Although the binding of

PqqA does not lead to large chemical shift perturbations in the α3 helix of PqqD, it is likely that there are interactions between these two elements. Interestingly, residues at the N-terminus of the α1 helix of PqqD show large chemical shift changes when complexed to PqqA (Figure 25). This could suggest that PqqA folds over the saddle between strand β2 and helix α2 of PqqD in a direct interaction, but the effect could also indirectly arise from the interaction between PqqA and strand β3 of PqqD, which is immediately N-terminal to helix α1.

PqqE binding to the PqqD–PqqA complex

For NisB, LynD, and MccB, the enzyme portion of the RRE-enzyme superstructure interacts with the RRE domain such that the β3–α1 and α2–α3 loops and the associated ends of those secondary structures are solvent exposed (Figure 26A–C). Results from the PqqD–PqqA–PqqE binding experiments indicate that the positioning of PqqD relative to PqqE is similar to the position seen in the RRE-enzyme superstructures of NisB, LynD, and MccB (Figure 26). RiPP precursor peptides are bipartite, with an N-terminal leader sequence involved in RRE-to-precursor peptide recognition, and a C-terminal core sequence that contains the residues ultimately comprising the final RiPP natural product.[11] Because the C-terminal core portion of PqqA must insert into the active site of PqqE, it is likely that the peptide is oriented such that the N-terminal end is close to the α2–α3 loop and the C-terminal end of the leader of PqqA lies at the β2–β3 loop, as observed in NisB and LynD. While the NisB and LynD leader sequences of the respective precursor peptides contain 23 and 43 residues, the peptide segments for which electron density was observed and modeled were only 13 and 15 residues long,

respectively, and represent only the residues critical to recognition (Figure 26E).[120, 121] In both cases, the amino acids bound to the RRE domain are significantly upstream from the start of the core peptide on which the enzyme acts (8 residues in NisB and 9 residues in LynD). In both cases, the active sites are relatively distant from the last ordered residue of the precursor peptide bound by the RRE (LynD, > 25 Å; NisB, > 30 Å). In contrast, the entire leader of PqqA is a similar length to the ordered peptide observed in the LynD and NisB crystal structures, with only 14 residues N-terminal to the Glu residue that is cross-linked to Tyr through the action of PqqE.[45] These observations suggest that the majority of the PqqA leader is likely bound to PqqD, placing the PqqA EXXXY core sequence very close to the RRE with few intervening residues such that the interaction between PqqD and the PqqE active site may be quite intimate. PqqE is a rSAM enzyme, and the chemistry must be sequestered from solvent.[61] Although the structure of PqqE is currently not known, an additional function of PqqD may be to act as a plug to prevent quenching of the radical by solvent during turnover.[61] This is in contrast to LynD and NisB, both dehydratases, where reaction byproducts need to exit the active site (ADP/phosphate and glutamate, respectively), and the RRE domains are more distant.[116, 120] This additional function may be specific to RiPPs whose biosynthesis involves rSAM enzymes, suggesting that even if the RRE is part of the polypeptide of the rSAM enzyme, it may require more mobility than the RRE domains of NisB, LynD, and MccB require. Supporting the hypothesis that PqqD acts as a plug is the presence of a long linker (26 residues) seen in the MePqqCD natural fusion (Figure 23).

It should be noted that the NMR experiments with PqqE were conducted aerobically, which means that the active site and SPASM domain Fe-S clusters were oxidized, and probably partially occupied; even the anaerobically isolated PqqE contains incomplete clusters that need to be reconstituted.[45] In addition to the [4Fe-4S] cluster of the rSAM active site, the PqqE SPASM domain contains a [4Fe-4S] and a second cluster that may be a [2Fe-2S].[45, 122] Electron paramagnetic resonance data indicate that the binding of PqqD affects at least one of the [4Fe-4S] clusters, which is consistent with the hypothesis that PqqD must bind in the proximity of the active site.[114]

It is becoming clear that rSAM-SPASM enzymes are widespread (~14000 annotations in the Interpro sequence database), particularly in association with RiPPs. The prevalence of rSAM enzymes in RiPP biosynthesis (both with and without SPASM domains) is likely linked to the inherent ability of radicals to initiate peptide chemistry. Bioinformatics suggests that ~50% of all RiPP rSAM-SPASM enzymes have an identifiable N-terminal RRE domain as part of the polypeptide, although in mycofactocin biosynthesis, like PQQ, the RRE is present as a separate protein, MftB.[44, 53, 54] The chemistry catalyzed by these enzymes includes carbon−carbon bond formation (PqqE[45] and StrB[123]), oxidative decarboxylation (MftC[53, 54]), and thioether bond formation (AlbA,[55, 56] SkfB,[57, 58] ThnB,[59] and SCIFF maturase[60]). The final RiPP products are chemically diverse with a wide range of physiological functions: antibiotic (thurincin H and subtilosin A), growth regulation (sporulation killing factor), signaling (streptide), and redox cofactors (PQQ). Unfortunately, the only crystal structure of a rSAM-SPASM enzyme is that of anSME, which is not part of a RiPP pathway but activates a sulfatase

under anaerobic conditions through post-translational modification.[62] Therefore, the structural details of how peptide substrates are presented to the prevalent rSAM-SPASM enzymes found within diverse RiPP biosynthesis pathways are currently unknown. This work provides the first insight by identifying residues in PqqD that interact with PqqE and showing that the binding surface overlaps with that of PqqA.

In conclusion, the first detailed monomeric structure of PqqD is presented. Mapping of chemical shifts perturbed by binding of PqqA, the precursor peptide, demonstrates that the binding mode of the PqqA leader peptide is likely similar to that observed in crystal structures of the unrelated dehydratases, LynD and NisB, utilizing a cleft formed between helix α3 and strand β3. The length of the ordered leader peptide bound to the RREs of LynD and NisB suggests that the core EXXXY residues of PqqA lie adjacent to the PqqD-bound peptide. This suggests that the PqqE active site lies very close to PqqD in the ternary complex, so PqqD may play the additional role of sequestering the PqqE radical intermediate from solvent. Structural work to precisely define these interactions is currently underway.

CHAPTER 4

This chapter is intended for future publication in *Acta Crystallographica Section F*

Crystal structure of *Methylobacterium extorquens* AM1 PqqC from the natural PqqCD fusion and the PqqC truncation

Robert L. Evans III, Morgan Esler, John A. Latham,

Judith P. Klinman, Carrie M. Wilmot

Paper 3 Abstract

Two crystal structures for PqqC, the final enzyme in the pyrroloquinoline quinone (PQQ) biosynthesis pathway, are reported. Both protein constructs are derived from *Methylobacterium extorquens* AM1 (Me) PqqCD (one, the full-length protein, and the other, the PqqC-domain truncation). MePqqCD and MePqqC constructs crystallized in space groups, $P4_12_12$ and $P2_12_12_1$, respectively, and the structures demonstrate that the MePqqC truncation has the native fold of the full-length protein. Additionally, MePqqCD and MePqqC are very similar to the published crystal structure of *Klebsiella pneumoniae* PqqC. However, the small PqqD domain of the MePqqCD crystal structure could not be modeled, as it forms no crystal contacts and is disordered in the solvent channels of the crystal.

INTRODUCTION

First characterized by J.G. Hauge in 1964, pyrroloquinoline quinone (PQQ) is a cofactor for prokaryotic aldose sugar and alcohol dehydrogenases.[1, 4-6, 15, 72] PQQ is synthesized independent of its affiliated dehydrogenase enzymes, functioning in these periplasmic dehydrogenases as an electron acceptor. This allows for the direct utilization of energy from applicable sugars and alcohols, and gives the bacteria a competitive growth advantage.[2] Interestingly, PQQ also enhances bacterial growth rates in species that do not synthesize the molecule.[6, 93, 94]

In addition to acting as a bacterial cofactor and vitamin, PQQ exhibits probiotic properties in eukaryotes, including: acting as a potent antioxidant;[31-33] contributing to mitochondriogenesis;[24-30] providing mammalian growth and reproduction benefits;[19-22] maintaining nervous tissue health;[124-127] and promoting plant growth.[73]

PQQ is classified as a ribosomally produced, post-translationally modified peptide (RiPP) biosynthesized by a comparatively simple post-ribosomal peptide synthesis (PRPS) pathway.[11] Just as PQQ is biochemically intriguing on its own, the biosynthetic pathway itself is also of broader interest as its simplicity makes it a model RiPP system for testing bioengineering strategies aimed at generating novel therapeutics based on other RiPP pathways.[11]

PQQ biosynthesis requires only five gene products from the *pqq* operon in *Klebsiella pneumoniae* (Kp): PqqA, the precursor peptide from which all nitrogens and carbons comprising PQQ originate; PqqB, a putative oxygenase; PqqC, an oxidase and the final enzyme in the pathway; PqqD, a scaffolding protein that presents PqqA to PqqE, a radical

S-adenosyl-methionine enzyme that initiates the pathway. A sixth protein, a putative protease PqqF, occurs in many species including Kp, but can be deleted with no phenotypic effects.[2, 44, 50, 97]

PqqC completes the synthesis of PQQ via four, two-electron transfer steps (eight electrons total), oxidizing and cyclizing AHQQ [3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid] to PQQ in the final step of the pathway (Figure 27).[47-50, 52]

PqqC is interesting because it activates oxygen without a metal nor redox prosthetic group (such as a flavin).[50] During the four two-electron cyclization/oxidation steps, either a dioxygen or a hydrogen peroxide molecule is reduced. In the case of O_2 as the electron acceptor, spin-forbidden, two-electron transfers from the substrate may be accomplished by two, single-electron transfers allowing time for spin-inversion to circumvent the formally spin-forbidden transfer.[128] This mechanism differs from how other oxidases that include redox cofactors activate dioxygen, as AHQQ itself plays dual roles of substrate and cofactor.[129-135] Therefore, PqqC is an intriguing member of a select group of oxidases classified as "cofactorless".[50] Understanding the diverse mechanisms employed in nature for controlled oxygen activation from the triplet to singlet state are critical to fundamentally understanding aerobic biology.

While PqqC and PqqD are naturally fused in *Methylobacterium extorquens* AM1 (Me), MePqqD expressed from a truncated Me*pqqCD* construct is natively folded and acts as a fully functional RiPP recognition element (RRE).[45, 136] Here we demonstrate that the MePqqC truncation is also natively folded when compared to the crystal structure of

the full-length MePqqCD, and we discuss the structure of MePqqC in relation to the previously published KpPqqC.

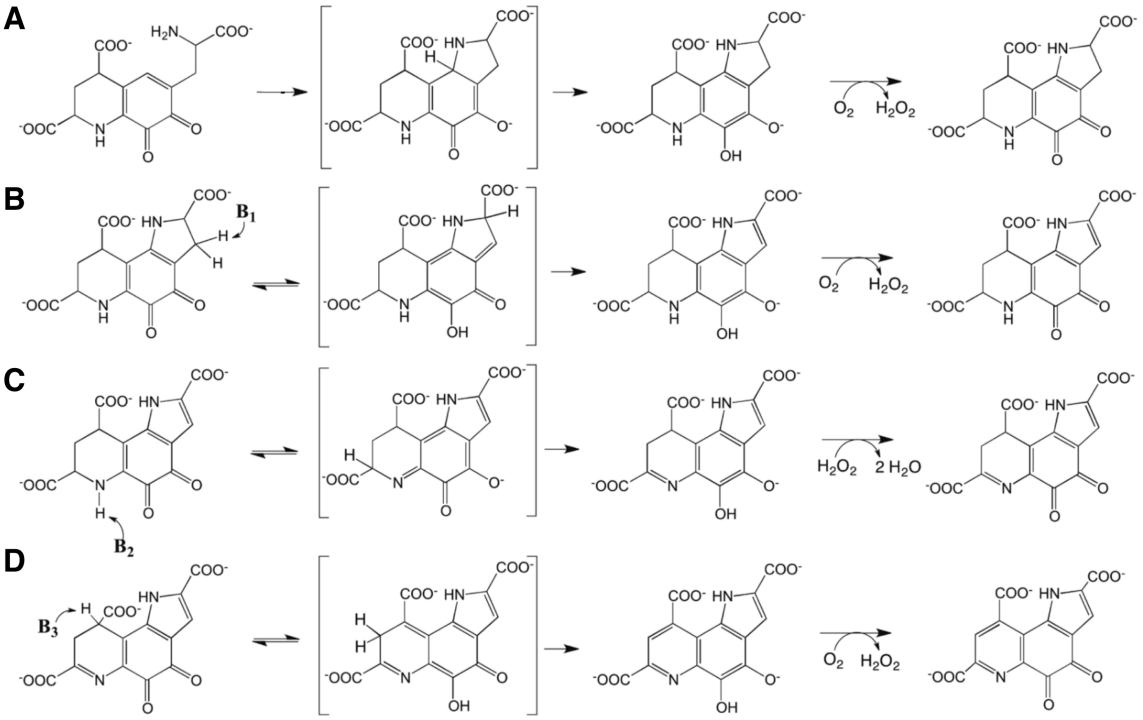

Figure 27. The substrate for PqqC, AHQQ, is converted to PQQ in a multistep, eight-electron cyclization (A) and oxidation (B, C, D). Dioxygen is the electron acceptor in steps A, B, and D. Hydrogen peroxide is the electron acceptor in step C. Graphic adapted from Scheme 1 of Bonnot, et al., (2013).[50]

MATERIALS AND METHODS

Cloning, gene expression and protein purification

Materials

The restriction enzymes used in all procedures came from New England BioLabs (Ipswich, MA), as did the T4 DNA ligase. All DNA constructs were sequenced by the University of California DNA Sequencing Facility (Berkeley, CA). Agilent (Santa Clara, CA) provided the *Pfu Turbo* polymerase, and Eurofins (Huntsville, AL) supplied

oligonucleotide primers. The pET28a vector was purchased from EMD Millipore (Bedford, MA). BL21 Gold (DE3) competent cells were from Agilent/Stratagene (Santa Clara, CA). PD-10 and HisTrap FF columns were products of GE Healthcare Life Sciences (Marlborough, MA)

Expression and purification of recombinant PqqCD

The N-terminal His_6-tagged *MepqqCD* (UniProt entry Q49150) gene (Figure 28A,B) was PCR amplified using *Methylobacterium extorquens* AM1 (ATCC, Manassas, VA) DNA as the template, purchased oligonucleotides as primers, and *Pfu Turbo* as the polymerase.

A

MTAQFPPPVPDTEQRLLSHEELEAALRDIGARRYHNLHPFHRLLHDGKLSKDQVRAWALN
RYYYQAMIPVKDAALLARLPDAQLRRIWRQRIVDHDGDHEGDGGIERWLKLAEGVGFTRD
YVLSTKGILSATRFSVDAYVHFVSERSLLEAIASSLTEMFSPTIISERVAGMLKNYDFIT
KDTLAYFDKRLTQAPRDADFALDYVKRHATTPEMQRAAIDALTFKCNVLWTQLDALYFAY
VAPGMVPPDAWQPGEGLVAETNSAEDSPAAAASPAATTAEPTAFSGSDVPRLPRGVRLRF
DEVRNKHVLLAPERTFDLDDNAVAVLKLVDGRNTVSQIAQILGQTYDADPAIIEADILPM
LAGLAQKRVLER

B

MGSSHHHHHHSSGLVPRGSHMTAQFPPPVPDTEQRLLSHEELEAALRDIGARRYHNLHPF
HRLLHDGKLSKDQVRAWALNRYYYQAMIPVKDAALLARLPDAQLRRIWRQRIVDHDGDHE
GDGGIERWLKLAEGVGFTRDYVLSTKGILSATRFSVDAYVHFVSERSLLEAIASSLTEMF
SPTIISERVAGMLKNYDFITKDTLAYFDKRLTQAPRDADFALDYVKRHATTPEMQRAAID
ALTFKCNVLWTQLDALYFAYVAPGMVPPDAWQPGEGLVAETNSAEDSPAAAASPAATTAE
PTAFSGSDVPRLPRGVRLRFDEVRNKHVLLAPERTFDLDDNAVAVLKLVDGRNTVSQIAQ
ILGQTYDADPAIIEADILPMLAGLAQKRVLER

C

MGSSHHHHHHSSGLVPRGSHMTAQFPPPVPDTEQRLLSHEELEAALRDIGARRYHNLHPF
HRLLHDGKLSKDQVRAWALNRYYYQAMIPVKDAALLARLPDAQLRRIWRQRIVDHDGDHE
GDGGIERWLKLAEGVGFTRDYVLSTKGILSATRFSVDAYVHFVSERSLLEAIASSLTEMF
SPTIISERVAGMLKNYDFITKDTLAYFDKRLTQAPRDADFALDYVKRHATTPEMQRAAID
ALTFKCNVLWTQLDALYFAYVAPGMVPPDAWQPGEGLVAETNSAED

Figure 28. (A) The primary sequence for MePqqCD natural fusion with the linker region colored green and the C-terminal PqqD domain in blue. (B) The recombinant

MePqqCD protein construct, with the N-terminal His_6-tag is red. (C) The MePqqC truncation, with the His_6-tag colored red and a small portion of the linker region colored green.

Products of the PCR amplification were digested using XhoI and NdeI and were subsequently purified by way of electrophoresis through agarose gel, after which the products were ligated into similarly-digested pET28a vectors using T4 DNA ligase. Sequence verification was performed on the cloned gene, which was then used to transform *Escherichia coli* BL21(DE3) cells. The *MePqqCD*/pET28a-transformed *E. coli* BL21(DE3) cells included an N-terminal His_6-tag for post-expression, IMAC purification. Cell grow-up, in LB media containing 50 μg/ml kanamycin, was maintained at 37 °C. Production of His_6-MePqqCD was induced by the addition of 1 mM isopropyl-D-galactopyranoside once the cell density had reached $A_{600} \sim 0.6$. TCEP (1 mM) was added to all buffers used in purification to prevent intermolecular disulfide bond formation. The induction period was 12 hours, during which temperature was maintained at 19 °C. Cells were harvested by 10 minutes of centrifugation at 6,500 rpm. Cells were resuspended in a buffer volume equaling five times the mass of cell paste. The buffer solution consisted of 50 mM imidazole, 50 mM Tris, pH 7.9, and 200 mM sodium chloride (lysis buffer). Cells were lysed using ultrasonication, and the lysate was centrifuged for 15 minutes at 20,000 rpm. The supernatant was loaded onto one 5-ml HisTrap FF column, which was then washed at 4 °C with lysis buffer to remove unwanted protein and then with elution buffer (consisting of 300 mM imidazole, 50 mM Tris, pH 7.9, and 200 mM sodium chloride) to elute the MePqqCD. The desired fractions were combined, concentrated, and buffer exchanged over PD-10 columns equilibrated

with 50 mM Tris, pH 7.9, 150 mM sodium chloride, and 1 mM TCEP. The combined sample was flash frozen in liquid nitrogen. SDS-PAGE analysis confirmed homogeneity. Final yield for the His_6-MePqqCD construct (Figure 28B) was 24 mg per liter of culture.

Expression and purification of recombinant PqqC

The DNA sequence encoding for the N-terminal His_6-tagged PqqC-domain truncation (residues 1-266, Figure 28C) of *M. extorquens* AM1 PqqCD (Uniprot entry Q49150) amino acid was amplified from genomic DNA (ATCC) and cloned into the *Nde*/*XhoI* restriction sites of pET28a. The sequence verified *pqqC/pET28a* plasmid was used to transform BL21 Gold (DE3) cells. An overnight culture was used to inoculate 4 L of LB broth and was then grown at 37° C to an $A_{600} \sim 0.6$. Upon reaching the indicated A_{600}, His_6-PqqC production was induced by the addition of 0.5 mM IPTG and the temperature was reduced to 20° C overnight. Cells from an overnight culture were harvested by centrifugation and suspended in lysis buffer (50 mM Tris, 200 mM sodium chloride, 30 mM imidazole, pH 7.9). Cells were lysed by sonication and clarified by centrifugation at 20,000 *g* for 10 minutes. The clarified lysate was applied to a 5 mL HisTrap FF column, after which the column was washed with 25 mL of lysis buffer. PqqC protein was eluted from the column with elution buffer (50 mM Tris, 200 mM sodium chloride, 300 mM imidazole, pH 7.9). Fractions containing purified PqqC as determined by SDS-PAGE, were pooled, concentrated, buffer exchanged into storage buffer (50 mM Tris, 100 mM sodium chloride, pH 7.9), and flash frozen in liquid nitrogen. Yield for the PqqC-domain protein construct was ~15 mg per liter of cell culture.

Screening, crystallization, and data collection

Crystallization screening was performed on a Rigaku CrystalMation robot. Each well contained 50 μL screening solution and each sitting drop contained 100 nL protein solution plus 100 nL well solution. Sealed plates were incubated at 20º C and the crystallization drops were digitally photographed using a Minstrel HT Imager. The generated images were inspected through a web interface provided on an internal CrystalTrak server. Candidate "hits" were optimized and screened for diffraction in-house on a Rigaku MSC Micromax 007 HFM X-ray generator equipped with a Saturn 944+ CCD Camera and an Oxford cryostream.

MePqqCD

Crystals were grown under hanging drop vapor diffusion conditions in 24-well VDX Plate trays with 500 μL crystallization volumes in the wells (Hampton Research). Drops were suspended from 22mm siliconized glass cover slips (Hampton Research). Protein solution was 8.0 mg/mL protein in 50 mM Tris, pH 7.9, 100 mM sodium chloride, and 1 mM TCEP. Well solution was 100 mM HEPES, pH 6.7, 19% w/v PEG-4000, and 10% isopropanol. The water used in the well solution contained 0.55 mM sodium azide for fungal growth suppression. The hanging drops were created by addition of 1 μL well solution to 1 μL protein solution. Trays incubated for 3 to 4 weeks before 0.5 μL of 0.08 mg/mL PQQ solution was added to each drop, after which, crystals appeared within 2 days. Oddly, crystallization would not commence without the addition of PQQ, but there was no evidence of PQQ binding in the resultant electron density. Crystals were generally needle-like, with the largest, best diffracting specimens reaching 20 μm × 20 μm × 400

μm. Before data collection, the crystal was transferred to a cryo-solution consisting of 25% glycerol and 75% well solution for ~30 s and was then flash-cooled in liquid nitrogen. Data were collected at 100 K on the Advanced Photon Source (APS) beamline 23-ID-D (GM/CA) of Argonne National Laboratory, Argonne, IL, USA using the *JBluIce-EPICS* software[137] and a Dectris Pilatus3-6m detector.

MePqqC truncation

The same hanging drop vapor diffusion set-up used for MePqqCD was used for the MePqqC truncation. Protein solution was 8.0 mg/mL protein in buffer containing 50 mM Tris, pH 8.0, and 200mM sodium chloride. Well solution was 100 mM HEPES, pH 8.1, 200 mM sodium chloride, and 23.75% w/v PEG-3350. All water used in well solution buffers contained 0.55 mM sodium azide. The protein solution to well solution ratio was 1:1 (1 μL to 1 μL). Unlike the MePqqCD trays, PQQ was not added to these drops. Crystals appeared after 1 to 3 weeks. The best diffracting crystal measured 170 μm × 290 μm. The crystal was cryoprotected in 25% glycerol and 75% well solution for ~30 s and flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K on APS beamline 23-ID-D using the *JBluIce-EPICS* software[137] and Dectris Pilatus3-6m detector.

Structure determination

Both data sets were integrated and scaled using *XDS*[138, 139] with integrated *JBluIce* pipeline automation developed by APS beamline 23-ID-D.

The 2.00 Å resolution dataset for MePqqC was phased first, and the resultant structure was then used to phase the 2.85 Å resolution MePqqCD dataset. MePqqC was

phased by molecular replacement with the single search model derived from the published KpPqqC coordinates (PDB entry 1OTV).[48] The search model was derived by manually editing away loop regions, and by pruning dissimilar side chains back to C_β atoms using a text editor and *Chainsaw*.[140] Molecular replacement was carried out using *Phaser*[141, 142] implemented in the *CCP4* suite.[143] A single solution, with four molecules in the asymmetric unit, was predominant, with RFZ = 3.3, TFZ = 20.4, PAK = 0, LLG = 1057. After several rounds of phase improvement and model building using *Refmac5*[143] and *Coot*,[144] and building only one of the polypeptide chains using the best monomer (chain A) electron density of the four monomers comprising the asymmetric unit (AU), a new search model was created and the original dataset was again processed with *Phaser*. The improved search model again gave a single solution with four molecules in the AU, but this time with an RFZ = 10.9, TFZ = 85.1, PAK = 0, and LLG = 14183. Following this, cycles of model building of all four chains of the AU and refinement were performed to yield the final model. *MolProbity*[110, 145] was used for structure validation and correction. The final run with *Refmac5* included TLS refinement.

The dataset for MePqqCD was phased by molecular replacement using an MePqqC monomer (from above) as the search model. As with the MePqqC crystal, the single solution provided by *Phaser* for MePqqCD had four molecules in the AU and an RFZ = 4.2, TFZ = 29.3, PAK = 0, and LLG = 4755. Refinement and validation were carried out as described above for MePqqC.

All figures were prepared with *MacPyMOL* (http://www.pymol.org).[113]

RESULTS AND DISCUSSION

Data collection statistics

The best diffracting crystal for both constructs yielded data collection and refinement statistics summarized in Table 3.

Table 4. Data collection and refinement statistics.

	MePqqCD	MePqqC trunc.
Data collection		
Space group:	$P4_12_12$	$P2_12_12_1$
Unit-cell parameters (Å, °):	a = 103.9 b = 103.9 c = 243.5, $\alpha = \beta = \gamma = 90$	a = 62.50 b = 114.2 c = 145.4, $\alpha = \beta = \gamma = 90$
Wavelength (Å):	1.0332	1.0332
Resolution (Å):	29.33–2.85 (2.95–2.85)*	29.51–2.00 (2.05–2.00)
R_{merge}:	0.062 (0.597)	0.077 (0.524)
$R_{measure}$:	0.070 (0.681)	0.083 (0.585)
$R_{p.i.m.}$:	0.032 (0.320)	0.031 (0.251)
Total No. of observations:	142,026	468,437
No. of unique reflections:	31,557	69,169
I/σ (I):	17.0 (2.1)	14.4 (1.9)
Completeness (%):	98.5 (91.5)	97.8 (67.7)
Multiplicity:	4.5 (4.2)	6.8 (4.6)
Wilson B factor (Å^2):	68.03	33.06
Source:	APS 23-IDD	APS 23-IDD
Refinement		
Resolution (Å):	2.85	2.00
No. of reflections:	29,899	65,610
R_{work}:	0.2531	0.1963
R_{free}[a]:	0.3271	0.2363
No. of protein residues:	815	910
No. of protein atoms:	6570	7010
Water molecules:	0	84
Average B factors (Å^2):	66.11	26.69
R.M.S. deviations…		
Bond lengths (Å):	0.0181	0.0277
Bond angles (°):	2.09	2.46
Ramachandran[b]…		
Favored (%):	90.5	98.5
Allowed (%):	7.4	1.3
Disallowed (%):	2.1	0.2
Clashscore[b]:	7.87	3.76
PDB ID:	5VRD	5VRC

*When used, values in parentheses are for the highest resolution shell.
[a]R_{free} is the R factor based on 5% of the data excluded from refinement.
[b]Value calculated by MolProbity (http://molprobity.biochem.duke.edu/).
Refinement was carried out in *Refmac* version 5.8.0158.

Phasing and refinement results

Refinement resulted in models with statistics as expected for the respective resolutions achieved, 2.85 Å for MePqqCD and 2.00 Å for MePqqC (Table 4). The final models both contain four molecules of PqqC in the AU. MePqqCD crystallized in space group $P4_12_12$ and MePqqC crystallized as $P2_12_12_1$. Although the MePqqCD protein construct included the PqqD-domain, no electron density was found and the domain could not be modeled.

Discussion comparing MePqqCD to the MePqqC truncation

No ordered water molecules were modeled for the lower resolution, MePqqCD, structure. A total of 84 waters were modeled into the higher resolution, MePqqC truncation, model. The crystallographic data for both proteins have been deposited in the Protein Data Bank (MePqqCD as PDB entry 5VRD and MePqqC as entry 5VRC).

The His_6-MePqqC truncation contained 288 residues (Figure 28C). Electron density for the 22-residue His_6-tag was not found for any of the 4 molecules in the AU indicating it was disordered and mobile. Protein residues that were modeled are as follows (residue numbering where 1 = the first residue of the actual gene product, not the His_6-tagged construct):

Chain A includes 228 residues: 14–162, 169–188, 200–258

Chain B includes 235 residues: 15–163, 166–190, 199–259

Chain C includes 227 residues: 15–162, 170–171, 173–189, 200–259

Chain D includes 224 residues: 16–157, 169–188, 198–259

His_6-MePqqCD contained 394 residues (Figure 28B). As in the truncation above, no electron density supporting the 22-residue His_6-tag was found, and further, no density supported residue modeling for the PqqC–D linker or the PqqD-domain, although ample solvent channel space at the C-terminal ends of the PqqC-domains would have allowed for the PqqD-domains (Figure 29). Protein residues that were modeled are as follows (again, residue numbering where 1 = the first residue of the gene product):

Chain A includes 210 residues: 15–98, 102–162, 173–179, 182–185, 199–252

Chain B includes 209 residues: 16–99, 102–162, 173–180, 197–252

Chain C includes 196 residues: 17–31, 33–98, 104–162, 175–179, 199–249

Chain D includes 204 residues: 19–31, 33–97, 102–162, 173–190, 199–245

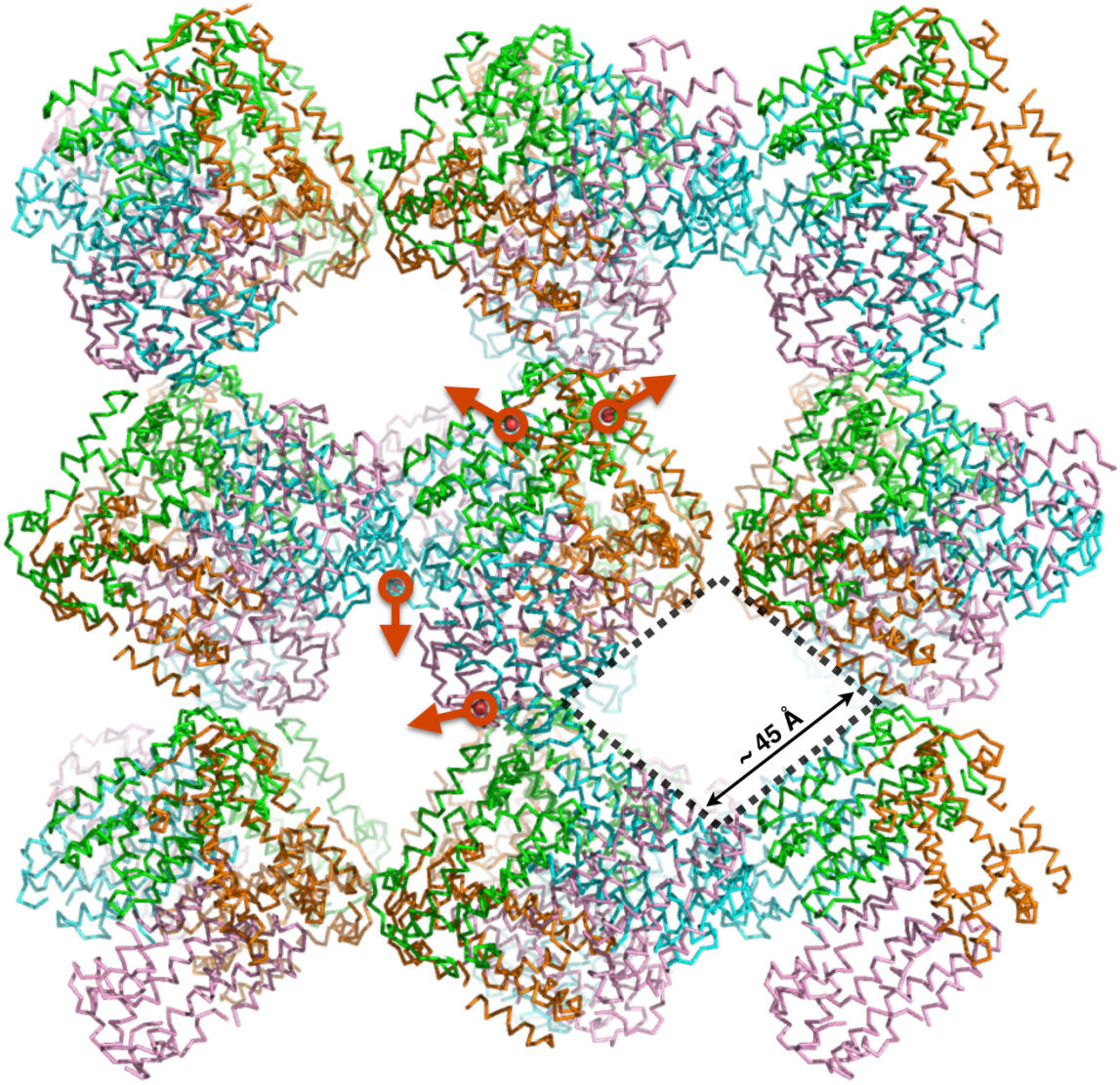

Figure 29. MePqqCD crystal packing. Rhomboidal solvent channels, ~ 45 Å to a side, run the length of the crystal, providing ample space for the 26-residue linker and 11 kDa PqqD-domain. The C-terminal ends for each of the four chains (red spheres) are all adjacent to these channels.

Average root mean square deviations (RMSDs) between the various chains for 2.85 Å MePqqCD (4 chains per AU), 2.00 Å MePqqC truncation (4 chains per AU), and 2.10 Å KpPqqC (PDB entry 1OTV, 2 chains per AU), were performed using the *PyMol* 'super' command, with cycles = 0. The RMSD alignments were performed using C-alphas. The

average RMSD between intramolecular chains in MePqqCD is 0.536 Å (population standard deviation (PSD 0.0972 Å)). The average RMSD for intramolecular chains in the MePqqC truncation is 0.730 Å (PSD 0.239 Å). When comparing all four chains from MePqqCD with all four chains from the MePqqC truncation, the RMSD is 0.927 Å (PSD 0.162 Å) demonstrating that the MePqqC structures are essentially the same (Figure 30). Both are in an "open" conformation where helices five and six are highly disordered, allowing for AHQQ/PQQ entry/departure (Figure 31A, gray helices). If the substrate was bound, the enzyme would have taken on a "closed" conformation, as seen in the KpPqqC-PQQ product complex (Figure 31B).[48]

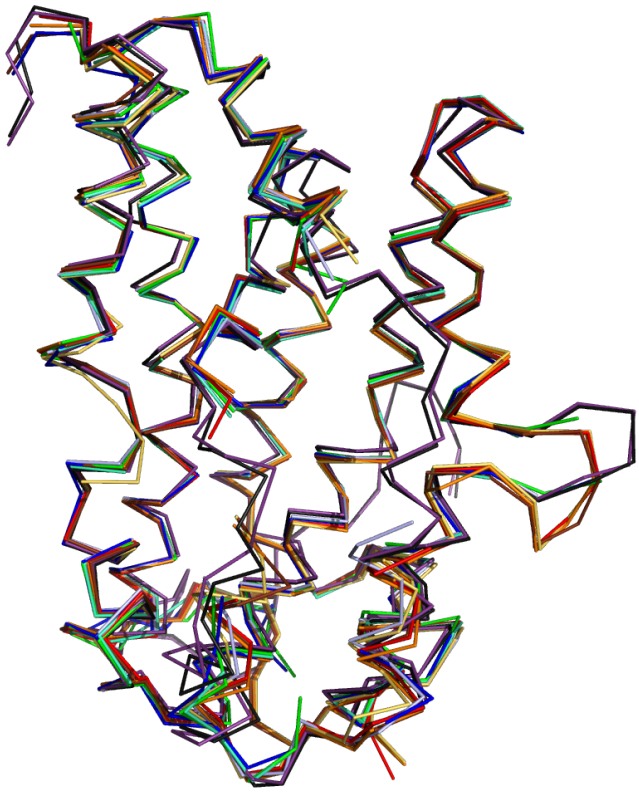

Figure 30. C-alpha traces of all four chains in the MePqqCD model (from A–D: green, green/cyan, light blue, dark blue), plus all four chains in the MePqqC truncation model (from A–D: brown, red, orange, yellow), plus the two chains in the

KpPqqC (PDB entry 1OTV) (chain A, violet; chain B, black). Alignments performed using the PyMol 'super' command.

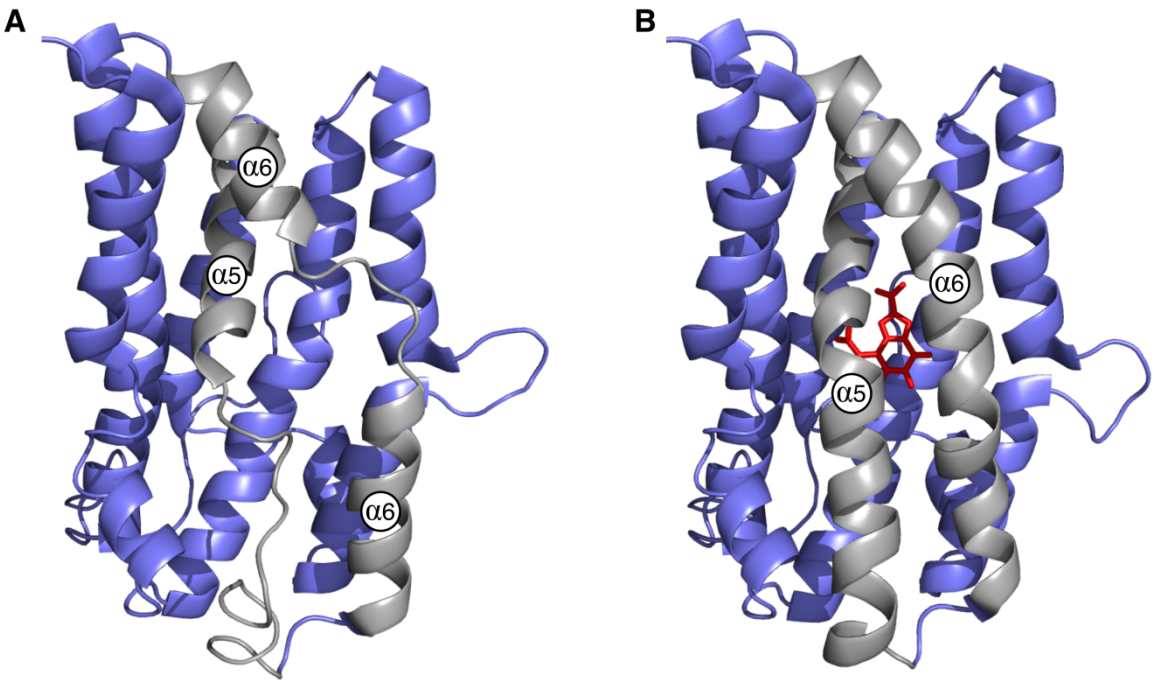

Figure 31. Illustrating the open and closed conformations of PqqC. The open (A) versus closed (B) conformations of PqqC can be seen in the deposited structures for KpPqqC and the KPPqqC-PQQ product complex (PDB entries 1OTV and 1OTW, respectively). The most significant differences between the conformations involve alpha-helices five and six (labeled and colored gray); in the absence of product (and presumably substrate) they are much more disordered. Once substrate binds, the helices become more ordered, although unpublished HDX mass spectrometry and fluorescence data from our collaborators in the Klinman lab, suggest that closure occurs progressively during the reaction. PQQ is colored red in panel B.

The tertiary structure for MePqqC and MePqqCD are in the open conformation (Figure 32). Both models show a seven helical bundle identical to the tertiary structure in KpPqqC.[48] Alpha-helix one (residues 19–46) is disrupted at residues 34–38. The longest helices in the open conformation, helices two and seven (residues 51–78 and 212–241, respectively), are continuous and highly ordered. Similar to helix one, helices three and

four (residues 82–115 and 119–145, respectively) are interrupted. Helix three, in particular, is interrupted by an interesting feature, a loop (residues 99–103) that appears MePqqC, MePqqCD, and in all deposited Kp structures. This highly conserved feature may have something to do with substrate binding, as it is adjacent to the helices that open and close to bind/release AHQQ/PQQ. Helices five and six are comprised of residues 147–175 and 181–208, respectively (Figure 32).

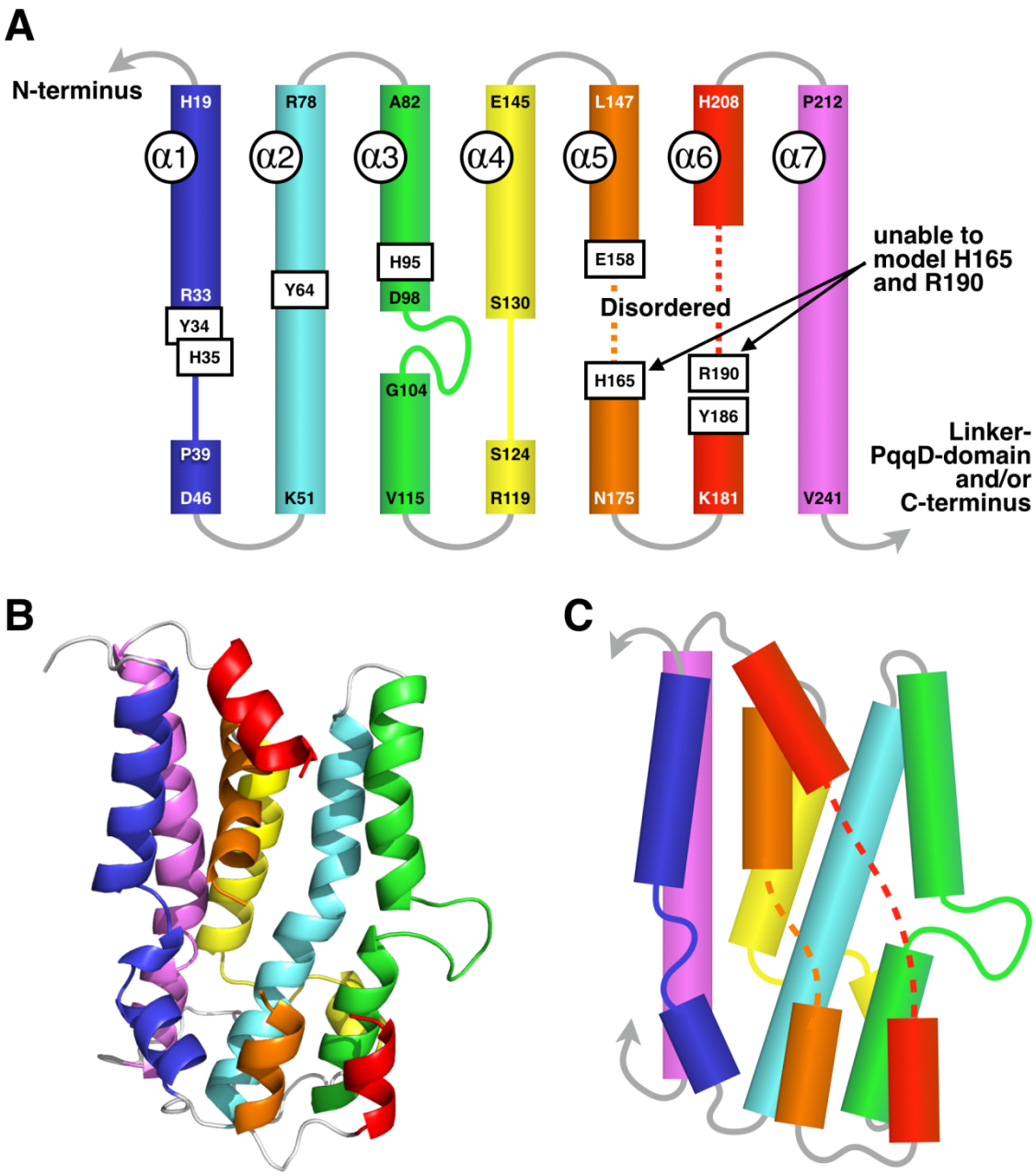

Figure 32. As in KpPqqC, MePqqC consists of seven alpha helices (A) arranged in a helical bundle (B, C). In the open conformation, illustrated here, helices five and six are disordered. The model shown in panel B is for the MePqqC truncation, although the PqqC-domain from MePqqCD is essentially identical. Active site residues (Y34, H35, Y64, H95, E158, H165, Y186, R190) are represented by boxes (panel A). Due to high disorder in helices five and six, active site residues H165 and R190 were not modeled.

The higher resolution MePqqC truncation structure included more C- and N-terminal residues and included fewer and smaller gaps (areas of weak density/high molecular mobility). The MePqqC truncation protein contained a few residues of the PqqC–D linker region, but only showed electron density for the first linker residue (Val 259). By contrast, the MePqqCD protein contained all residues out to and including the PqqD-domain when the crystallizations were set up, and yet only had electron density up to Trp 252. There is the possibility that the MePqqCD underwent proteolysis in the hanging drop prior to crystal formation, cleaving away part or all of the linker—PqqD-domain, and only this truncated protein crystallized. Before submitting this dissertation chapter for publication, an SDS-PAGE gel of dissolved crystals from the experimental tray will be run to see if the form that crystallized is truncated or appears full-length. If a single PqqCD band is observed, the lack of electron density can be attributed to a highly disordered linker—PqqD-domain in the large solvent channels (Figure 29).

Ramachandran analysis showed that the lower resolution MePqqCD structure has poorer model statistics, as expected. Only 90.5% (738/815) of all residues were in favored regions, and 97.9% (798/815) of all residues were in allowed regions. In comparison, for MePqqC, 98.5% (875/888) of all residues were in favored regions, and 99.8% (886/888) of all residues were in allowed regions.

Discussion comparing the MePqqC (truncation) to KpPqqC

The 2.0 Å MePqqC truncation structure is clearly superior to the structure for the 2.85 Å MePqqCD natural fusion, therefore comparisons to the deposited KpPqqC structure (PDB entry 1OTV at 2.1 Å resolution)[48] will only be made with MePqqC. Sequence

identity between KpPqqC and MePqqC is 41.9% (Clustal Omega (1.2.4), www.ebi.ac.uk). The average RMSD on C-alphas for all four chains from MePqqC, compared to the two chains from KpPqqC, is 1.71 Å (PSD 0.293 Å). The 2*F*o − *F*c map for three of the active site residues that are conserved between MePqqC and KpPqqC[48, 50] illustrates the quality of the electron density (Figure 33).

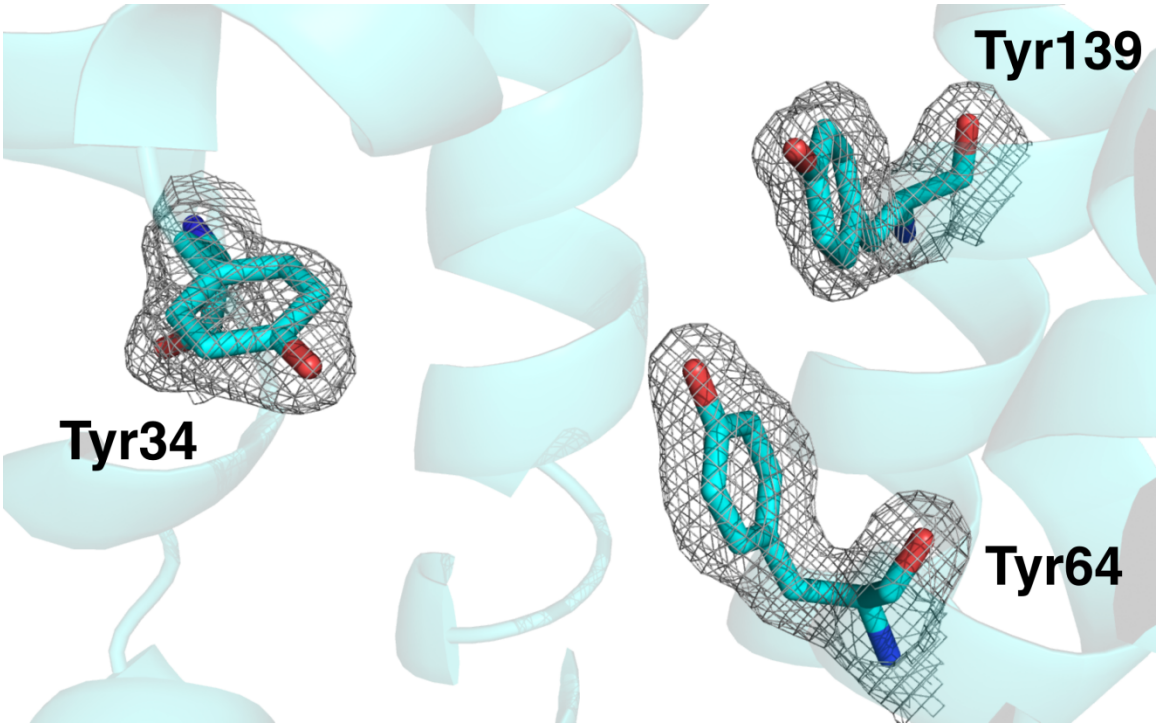

Figure 33. MePqqC 2Fo - Fc density for three active site residues. Density was contoured at sigma = 1.0, carve = 2.5. The map was calculated using *FFT*.[146]

The MePqqC crystal structure confirms the presence and positioning of previously identified active site residues in KpPqqC.[48, 50, 147] MePqqC shows similarly arranged residues at the active site (Figure 34, circled residues).

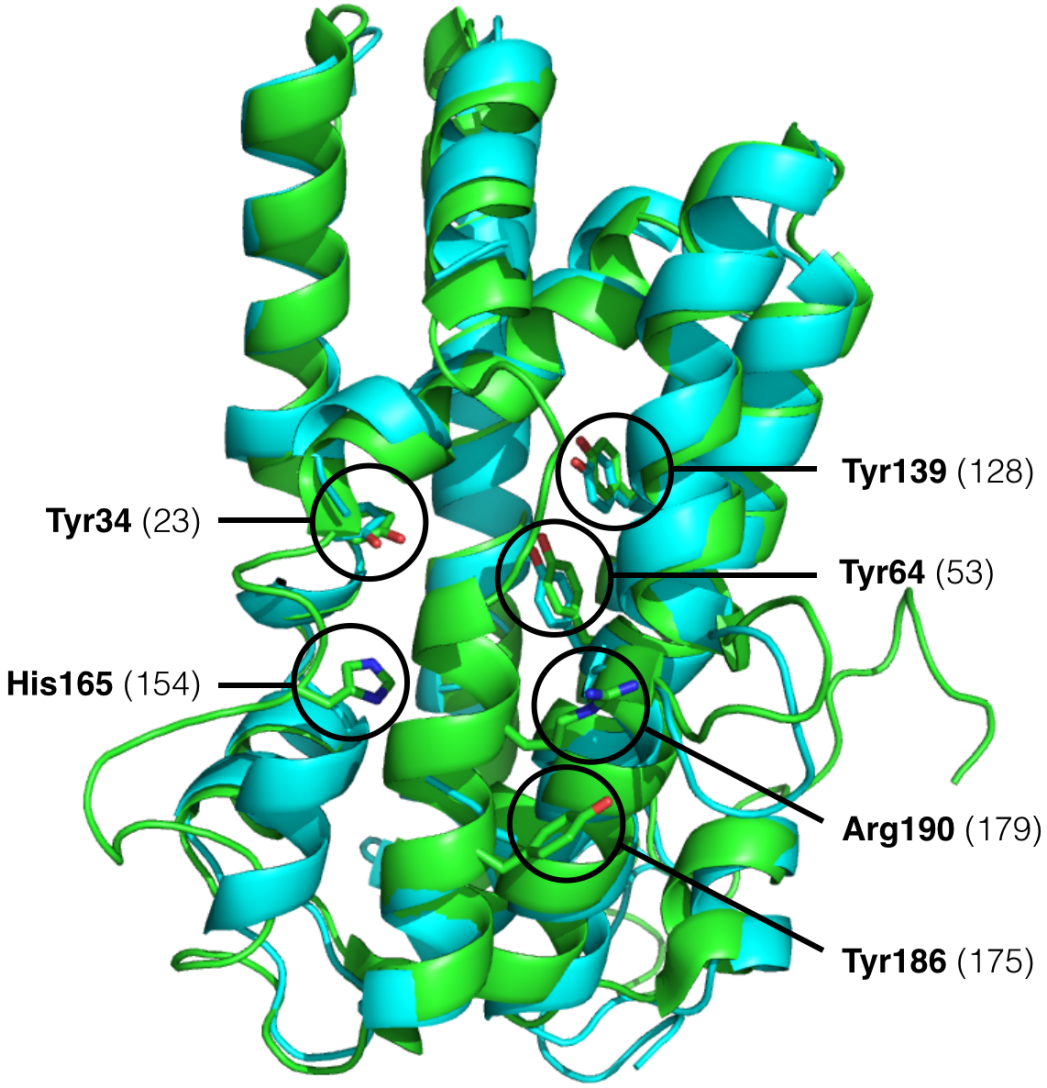

Figure 34. Cartoon structure of MePqqC truncation (cyan) overlaid with KpPqqC (green). Active site residues are circled. The helices that contain the active site residues His165, Arg190, and Tyr186 are even more disordered in MePqqC than in KpPqqC, therefore, these three residues could not be modeled for MePqqC.

CONCLUSIONS

Although PQQ was required to crystallize MePqqCD, no electron density supporting the presence of the molecule was observed (Figure 33) and the enzyme was in an "open" conformation (Figures 31A, 32B, 34 cyan). This inability (under these crystal conditions) to bind PQQ stands in contrast to what has been observed with KpPqqC, where co-crystallization has proven viable. Electron density supporting PQQ and a putative H_2O_2 product has been observed at the active site of KpPqqC (PDB entry 1OTW)[48] and evidence of bound $PQQH_2$ was seen in the KpPqqC Y175F mutant co-crystallized with AHQQ (PDB entry 4NY7).[148] Further, attempts to soak PQQ into the MePqqC truncation crystals did not lead to PQQ binding either, and when PQQ was added in excess, the MePqqC truncation crystals rapidly dissolved (~30 s).

The MePqqC substrate entry helices (helices 5 and 6) appear to be accessible through solvent channels and do not appear to be involved in crystal contacts (Figure 29), but they do show a higher level of disorder (a lack of electron density and fewer modeled residues) than what is seen in KpPqqC (Figure 34) even though the two structures are of similar resolutions (KpPqqC is 2.1 Å and MePqqC is 2.0 Å). Therefore, there would be a higher entropic cost to binding PQQ. In the MePqqC crystals, addition of higher concentrations of PQQ led to the crystals dissolving, suggesting that the crystal packing could not survive the disorder-to-order transition that occurs upon PQQ binding.

Regardless of our ability to capture PQQ in the active site, there are now two additional PqqC structures from another organism deposited in the PDB, entry 5VRC for MePqqC and entry 5VRD for MePqqCD.

CHAPTER 5

Conclusions

Accomplishments

NMR accomplishments (PqqD-related)

Success in solving the solution NMR structure for the MePqqD truncation means that the first detailed monomeric structure of PqqD (physiological form) is now available in the PDB (entry 5SXY). More than that, this structure represents the first *independent* RRE to be deposited to the PDB and is, even further, the first structure of an RRE that is involved with a PRPS rSAM-SPASM enzyme.

The binding studies for the PqqD–PqqA and PqqD–PqqA–PqqE complexes, show that the majority of the PqqA leader is likely bound to PqqD, placing the PqqA EXXXY core sequence very close to the RRE with few intervening residues. This proximal positioning indicates that the interaction between PqqD and the PqqE active site may be quite close and critical. PqqE is a radical-generating rSAM enzyme,[45] and in addition to PqqD's RRE roles of peptide recognition, binding, and presentation,[11, 42, 44] an additional function of PqqD may be to shield the tailoring enzyme's active site chemistry from radical-quenching solvent.

XRC accomplishments (PqqC-related)

While we were eventually able to crystalize both MePqqCD and the MePqqC truncation, neither of these crystals took up PQQ in soak experiments, making them unsuitable for trapping intermediates via anaerobic soaks. However, they could be useful

in anaerobic co-crystallization experiments with AHQQ, and I suggest this below. So, while neither crystal form showed promise for soaking experiments, we did contribute two new structures to the PDB (entry 5VRC represents the MePqqC truncation and entry 5VRD represents MePqqCD, both entries are on hold pending publication of paper 3 (chapter 4 of this dissertation)).

XRC accomplishments (PqqB-related)

PqqB crystallographic studies were ultimately transferred to Chao Li, in the Wilmot Lab, but prior to that transfer, I was able to solve a structure for the *Pseudomonas putida* PqqB homodimer complexed with zinc (active site physiological metal is thought to be copper) and a pre-AHQQ mimic, 5-S-cysteinyl-DOPA (Figure 5). The true pre-AHQQ molecule is expected to contain three carboxyl moieties, rather than the two contained in 5-S-cysteinyl-DOPA mimic. The lack of the additional carboxylate group in the mimic led to ambiguous electron density that supported multiple binding conformations. Chao and his collaborators in the Klinman Lab, are pursuing better pre-AHQQ mimics.

Other significant XRC accomplishments (a project not included in this dissertation)

In a collaboration with Bryan Jones (Kazlauskas Lab, BMBB, UMN), I assisted in solving the crystal structure of a possible ancestral hydroxynitrile lyase derived by extrapolation back from two modern, plant enzymes belonging to the α/β-hydrolase superfamily. My work included guiding Bryan through the process of crystal growth screening, crystal optimization, cryoprotection, harvesting/mounting, data collection both in-house and at the synchrotron, data processing, phasing, refinement, and deposition. Publication of this work is underway.

Summary of structures

During my time as a graduate student in the Wilmot Lab, I either directly or indirectly, solved numerous structures, four that are currently deposited in the RCSB PDB, and three that are works in-progress (Table 5).

Table 5. Summary of structures that I either solved or for which I made significant contributions

PDB Entry	Method	Description	Status	Publication
5SXY	NMR	MePqqD truncation 1st physiological PqqD, 1st stand-alone RRE	available	*Biochemistry*. 2017 May 30;56(21):2735-2746. DOI: 10.1021/acs.biochem.7b00247
5TDX	XRC	putative ancestral hydroxynitrile lyase, containing an α/β-hydrolase fold (cladistically-guided design, plant kingdom)	pending publication	Journal: TBD Working title: Echoes of the Past: Structure and evolutionary insights from the reconstructed ancestral plant hydroxynitrile lyase, HNL1
5VRC	XRC	MePqqCD	pending publication	Journal: *Acta Crystallographica Section F* Working title: Crystal structure of *Methylobacterium extorquens* PqqC from the natural fusion and the truncation
5VRD	XRC	MePqqC truncation	pending publication	Journal: *Acta Crystallographica Section F* Working title: Crystal structure of *Methylobacterium extorquens* PqqC from the natural fusion and the truncation
—	XRC	MePqqD truncation, SeMet $P2_12_12_1$ domain-swapped dimer (phased via Se-SAD)	work in-progress	Journal: TBD Working title: TBD
—	NMR	MePqqD–PqqA binary complex	work in-progress	Journal: TBD Working title: TBD
—	XRC	KpPqqC (true wild type) new space group *I*222 vs. $P2_12_12$ for A21D "WT"	work in-progress	Journal: TBD Working title: TBD

Future Directions

Experiments to determine PqqA residues involved in binding

HSQC titrations using specific peptide leader sequence mutations could elucidate RRE–peptide (PqqD–PqqA) recognition and binding permissiveness, ultimately informing bioengineering efforts.[11] Chemical shifts (peak migrations) from these HSQC titration experiments could be used to calculate K_D values for each construct, provided

binding is not too tight (where K_D values are ≥ ~3 μM).[149] Surface plasmon resonance (SPR) spectroscopy would also be useful for examining association/dissociation kinetic values (k_{on} and k_{off})for the binary complex.[150] NMR HSQC experiments, in comparison, provide information on the relative involvements of specific residues in binding (greater weighted peak shift changes indicate more significant changes in individual chemical environments for residues). Isothermal titration calorimetry (ITC) can also provide binding information (e.g. the dissociation constant, K_D), but, like SPR, provides overall information about peptide construct binding, rather than information about the relative interactions of specific peptide residues.[150] ITC and SPR have the advantage of not requiring isotopically-labeled PqqD (or PqqA) and, therefore, cost less.

NMR experiments to solve binary structure

We know the orientation of PqqD to PqqE in the PqqD–PqqA–PqqE complex (Figures 25 and 26D), and from this information, can infer the orientation of PqqA with respect to PqqD (the C-terminal core portion of the peptide must be inserted into the enzyme for modification, and therefore can only be oriented one way). However, knowing the relative positioning of PqqA when bound to PqqD is not the same as having a complete structure showing detailed positioning for all residues on both the protein and the peptide. Many of the through-bond experiments leading to peak assignment for PqqD in the PqqD–PqqA complex, have been completed. The same type of experiments will need to be completed for ^{13}C,^{15}N-labeled-PqqA in the PqqD–PqqA complex, in order to assign all ^{1}H, ^{13}C, ^{15}N, and peaks for the peptide. Further, NOE (through space) experiments will need to be completed for the labeled-PqqD/unlabeled-PqqA, unlabeled-

PqqD/labeled-PqqA, and isotope-edited labeled-PqqD/labeled-PqqA binary constructs. This is a substantial undertaking, every bit as challenging as the work needed for solving the stand-alone PqqD structure. The resulting structure will give more details of RRE–peptide-enzyme interactions. Finally, experiments to define the PqqA residues involved in binding will also require PqqE; unlabeled PqqD/labeled PqqA spectra should be compared with the unlabeled PqqD/labeled PqqA/unlabeled PqqE spectra to completely understand peptide binding dynamics in the contexts of both binary and ternary complexation. These experiments will not require labeled PqqE, nor will experiments with PqqE include NOEs, as a structure of the ternary complex is not the goal here.

Crystallography experiments to solve PqqD structure and the binary complex

Ironically, after failing to find crystallization conditions for PqqD or PqqD–PqqA, and then turning to NMR in order to solve the PqqD structure, crystallizing conditions for MePqqD were serendipitously discovered while purifying the protein for NMR work. MePqqD protein was observed to spontaneously crystallize into thin plates when buffer-exchanged into pure 25 mM potassium phosphate (pH 6.5) and stored at 4º C. Without any further crystal optimization, a crystal harvested from the side of a 50-mL conical tube, cleanly diffracted out to 1.72 Å. Multiple data sets were collected, but none were able to be phased using molecular replacement (using either the deposited XcPqqD domain-swapped dimer (3G2B)[63] or the MePqqD solution NMR structure (5SXY)[136] for search model development). Recently, after over expressing the MePqqD protein using a Se-methionine (Se-Met) protocol, Morgan Esler (Wilmot Lab) acquired a 1.96 Å dataset collected just above the Se edge, allowing for Se-Met SAD phasing. The resulting

structure is currently being used as a molecular replacement search model for the higher-resolution (1.72 Å) dataset that we were formerly unable to solve. We now know that this new MePqqD construct crystallized via a domain swap, as did XcPqqD.[63] This knowledge will help us understand more about PqqD dynamics and could aid in developing strategies for crystallization aimed at binary and ternary complex structures.

Crystallography experiments to solve ternary complex

No PqqE structure has yet been deposited in the PDB, although collaborators at the Albert Einstein College of Medicine (Bronx, NY) have come close to completing a 3.2 Å structure for *M. extorquens* AM1 PqqE, although it has lost the [4Fe-4S] cluster at the rSAM active site. Successes in solving pieces of this complex, as described above, will help in phasing datasets collected for a future ternary complex. Eventually, we would like to see completed ternary complex structures for not only PqqD–PqqA–PqqE, but for other examples of RiPP-initiating RRE–peptide–rSAM-SPASM complexes.

Crystallography experiments to trap AHQQ-to-PQQ intermediates

While the crystal forms for MePqqC/CD were unable to bind PQQ in soak experiments while retaining crystal integrity, anaerobic *co-crystallization* of MePqqC/CD with AHQQ as the starting point may allow trapping of reaction intermediates through introduction of small amounts of oxygen, as was done to trap catalytic intermediates in crystals of copper amine oxidase in the Wilmot Lab.[151]

Now that we know that the "wild type" KpPqqC structure, for which we were unable to replicate crystal growth conditions, was actually an A21D mutant, we can use the mutant construct to grow crystals and pursue intermediate trapping as above or through

soaking with AHQQ. Single-crystal UV-Vis spectroscopy can be used to corroborate the presence of specific intermediates.[152, 153]

References

[1] Hauge, J. G. (1964) Glucose Dehydrogenase of Bacterium Anitratum: An Enzyme with a Novel Prosthetic Group, *J. Biol. Chem. 239*, 3630-3639.

[2] Klinman, J. P., and Bonnot, F. (2014) Intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ, *Chem. Rev. 114*, 4343-4365.

[3] Okeley, N. M., and van der Donk, W. A. (2000) Novel cofactors via post-translational modifications of enzyme active sites, *Chem. Biol. 7*, R159-171.

[4] Meulenberg, J. J., Sellink, E., Loenen, W. A., Riegman, N. H., van Kleef, M., and Postma, P. W. (1990) Cloning of Klebsiella pneumoniae pqq genes and PQQ biosynthesis in Escherichia coli, *FEMS Microbiol. Lett. 59*, 337-343.

[5] Meulenberg, J. J., Sellink, E., Riegman, N. H., and Postma, P. W. (1992) Nucleotide sequence and structure of the Klebsiella pneumoniae pqq operon, *Mol. Gen. Genet.: MGG 232*, 284-294.

[6] Velterop, J. S., Sellink, E., Meulenberg, J. J. M., David, S., Bulder, I., and Postma, P. W. (1995) Synthesis of Pyrroloquinoline Quinone in-Vivo and in-Vitro and Detection of an Intermediate in the Biosynthetic-Pathway, *J. Bacteriol. 177*, 5088-5098.

[7] Janes, S. M., Mu, D., Wemmer, D., Smith, A. J., Kaur, S., Maltby, D., Burlingame, A. L., and Klinman, J. P. (1990) A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of bovine serum amine oxidase, *Science 248*, 981-987.

[8] McIntire, W. S., Wemmer, D. E., Chistoserdov, A., and Lidstrom, M. E. (1991) A new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as the redox prosthetic group in methylamine dehydrogenase, *Science 252*, 817-824.

[9] Wang, S. X., Mure, M., Medzihradszky, K. F., Burlingame, A. L., Brown, D. E., Dooley, D. M., Smith, A. J., Kagan, H. M., and Klinman, J. P. (1996) A crosslinked cofactor in lysyl oxidase: redox function for amino acid side chains, *Science 273*, 1078-1084.

[10] Datta, S., Mori, Y., Takagi, K., Kawaguchi, K., Chen, Z. W., Okajima, T., Kuroda, S., Ikeda, T., Kano, K., Tanizawa, K., and Mathews, F. S. (2001) Structure of a quinohemoprotein amine dehydrogenase with an uncommon redox cofactor and highly unusual crosslinking, *Proc. Natl. Acad. Sci. U.S.A. 98*, 14268-14273.

[11] Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C.,

Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep. 30*, 108-160.

[12] Shen, Y. Q., Bonnot, F., Imsand, E. M., RoseFigura, J. M., Sjolander, K., and Klinman, J. P. (2012) Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone, *Biochemistry 51*, 2265-2275.

[13] McCarthy, M. (2017) Woman dies after infection with bacteria resistant to all antibiotics available in US, *BMJ 356*, j254.

[14] Toyama, H., Chistoserdova, L., and Lidstrom, M. E. (1997) Sequence analysis of pqq genes required for biosynthesis of pyrroloquinoline quinone in Methylobacterium extorquens AM1 and the purification of a biosynthetic intermediate, *Microbiology 143 (Pt 2)*, 595-602.

[15] Goodwin, P. M., and Anthony, C. (1998) The biochemistry, physiology and genetics of PQQ and PQQ-containing enzymes, *Adv. Microb. Physiol. 40*, 1-80.

[16] Duine, J. A. (1999) The PQQ story, *J. Biosci. Bioeng. 88*, 231-236.

[17] Anthony, C. (2004) The quinoprotein dehydrogenases for methanol and glucose, *Arch. Biochem. Biophys. 428*, 2-9.

[18] Matsumura, H., Umezawa, K., Takeda, K., Sugimoto, N., Ishida, T., Samejima, M., Ohno, H., Yoshida, M., Igarashi, K., and Nakamura, N. (2014) Discovery of a eukaryotic pyrroloquinoline quinone-dependent oxidoreductase belonging to a

new auxiliary activity family in the database of carbohydrate-active enzymes, *PLoS one 9*, e104851.

[19] Kasahara, T., and Kato, T. (2003) Nutritional biochemistry: A new redox-cofactor vitamin for mammals, *Nature 422*, 832.

[20] Killgore, J., Smidt, C., Duich, L., Romero-Chapman, N., Tinker, D., Reiser, K., Melko, M., Hyde, D., and Rucker, R. B. (1989) Nutritional importance of pyrroloquinoline quinone, *Science 245*, 850-852.

[21] Steinberg, F., Stites, T. E., Anderson, P., Storms, D., Chan, I., Eghbali, S., and Rucker, R. (2003) Pyrroloquinoline quinone improves growth and reproductive performance in mice fed chemically defined diets, *Exp. Biol. Med. (Maywood) 228*, 160-166.

[22] Steinberg, F. M., Gershwin, M. E., and Rucker, R. B. (1994) Dietary pyrroloquinoline quinone: growth and immune response in BALB/c mice, *J. Nutr. 124*, 744-753.

[23] Ohwada, K., Takeda, H., Yamazaki, M., Isogai, H., Nakano, M., Shimomura, M., Fukui, K., and Urano, S. (2008) Pyrroloquinoline Quinone (PQQ) Prevents Cognitive Deficit Caused by Oxidative Stress in Rats, *J. Clin. Biochem. Nutr. 42*, 29-34.

[24] Bauerly, K., Harris, C., Chowanadisai, W., Graham, J., Havel, P. J., Tchaparian, E., Satre, M., Karliner, J. S., and Rucker, R. B. (2011) Altering pyrroloquinoline quinone nutritional status modulates mitochondrial, lipid, and energy metabolism in rats, *PLoS one 6*, e21779.

[25] Bauerly, K. A., Storms, D. H., Harris, C. B., Hajizadeh, S., Sun, M. Y., Cheung, C. P., Satre, M. A., Fascetti, A. J., Tchaparian, E., and Rucker, R. B. (2006) Pyrroloquinoline quinone nutritional status alters lysine metabolism and modulates mitochondrial DNA content in the mouse and rat, *Biochim. Biophys. Acta 1760*, 1741-1748.

[26] Chowanadisai, W., Bauerly, K. A., Tchaparian, E., Wong, A., Cortopassi, G. A., and Rucker, R. B. (2010) Pyrroloquinoline quinone stimulates mitochondrial biogenesis through cAMP response element-binding protein phosphorylation and increased PGC-1alpha expression, *J. Biol. Chem. 285*, 142-152.

[27] Harris, C. B., Chowanadisai, W., Mishchuk, D. O., Satre, M. A., Slupsky, C. M., and Rucker, R. B. (2013) Dietary pyrroloquinoline quinone (PQQ) alters indicators of inflammation and mitochondrial-related metabolism in human subjects, *J. Nutr. Biochem. 24*, 2076-2084.

[28] Stites, T., Storms, D., Bauerly, K., Mah, J., Harris, C., Fascetti, A., Rogers, Q., Tchaparian, E., Satre, M., and Rucker, R. B. (2006) Pyrroloquinoline quinone modulates mitochondrial quantity and function in mice, *J. Nutr. 136*, 390-396.

[29] Singh, A. K., Pandey, S. K., Saha, G., and Gattupalli, N. K. (2015) Pyrroloquinoline quinone (PQQ) producing Escherichia coli Nissle 1917 (EcN) alleviates age associated oxidative stress and hyperlipidemia, and improves mitochondrial function in ageing rats, *Exp. Gerontol. 66*, 1-9.

[30] Zhang, J., Meruvu, S., Bedi, Y. S., Chau, J., Arguelles, A., Rucker, R., and Choudhury, M. (2015) Pyrroloquinoline quinone increases the expression and activity of Sirt1 and -3 genes in HepG2 cells, *Nutr. Res. 35*, 844-849.

[31] Fluckiger, R., Paz, M., Mah, J., Bishop, A., and Gallop, P. M. (1993) Characterization of the glycine-dependent redox-cycling activity in animal fluids and tissues using specific inhibitors and activators: evidence for presence of PQQ, *Biochem. Biophys. Res. Commun. 196*, 61-68.

[32] Fluckiger, R., Paz, M. A., and Gallop, P. M. (1995) Redox-cycling detection of dialyzable pyrroloquinoline quinone and quinoproteins, *Methods Enzymol. 258*, 140-149.

[33] Stites, T. E., Mitchell, A. E., and Rucker, R. B. (2000) Physiological Importance of Quinoenzymes and the O-Quinone Family of Cofactors, *J. Nutr. 130*, 719-727.

[34] Marahiel, M. A. (2009) Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis, *J. Peptide Science 15*, 799-807.

[35] Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J., and Rebuffat, S. (2007) Microcins, gene-encoded antibacterial peptides from enterobacteria, *Nat. Prod. Rep. 24*, 708-734.

[36] Frey, P. A., Hegeman, A. D., and Ruzicka, F. J. (2008) The Radical SAM Superfamily, *Crit. Rev. Biochem. Mol. Biol. 43*, 63-88.

[37] Maksimov, M. O., Pan, S. J., and James Link, A. (2012) Lasso peptides: structure, function, biosynthesis, and engineering, *Nat. Prod. Rep. 29*, 996-1006.

[38] Bushnell, D. A., Cramer, P., and Kornberg, R. D. (2002) Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution, *Proc. Natl. Acad. Sci. U.S.A. 99*, 1218-1222.

[39] Wieland, T., and Govindan, V. M. (1974) Phallotoxins bind to actins, *FEBS Lett. 46*, 351-353.

[40] Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif, *J. Mol. Biol. 294*, 1327-1336.

[41] Buczek, O., Bulaj, G., and Olivera, B. M. (2005) Conotoxins and the posttranslational modification of secreted gene products, *Cell. Mol. Life Sci. 62*, 3067-3079.

[42] Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis, *Nat. Chem. Biol. 11*, 564-570.

[43] Solbiati, J. O., Ciaccio, M., Farias, R. N., Gonzalez-Pastor, J. E., Moreno, F., and Salomon, R. A. (1999) Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25, *J. Bacteriol. 181*, 2659-2662.

[44] Latham, J. A., Iavarone, A. T., Barr, I., Juthani, P. V., and Klinman, J. P. (2015) PqqD is a novel peptide chaperone that forms a ternary complex with the radical S-adenosylmethionine protein PqqE in the pyrroloquinoline quinone biosynthetic pathway, *J. Biol. Chem. 290*, 12908-12918.

[45] Barr, I., Latham, J. A., Iavarone, A. T., Chantarojsiri, T., Hwang, J. D., and Klinman, J. P. (2016) Demonstration That the Radical S-Adenosylmethionine (SAM) Enzyme PqqE Catalyzes de Novo Carbon-Carbon Cross-linking within a Peptide Substrate PqqA in the Presence of the Peptide Chaperone PqqD, *J. Biol. Chem. 291*, 8877-8884.

[46] Podzelinska, K., He, S. M., Wathier, M., Yakunin, A., Proudfoot, M., Hove-Jensen, B., Zechel, D. L., and Jia, Z. (2009) Structure of PhnP, a phosphodiesterase of the carbon-phosphorus lyase pathway for phosphonate degradation, *J. Biol. Chem. 284*, 17216-17226.

[47] Toyama, H., Fukumoto, H., Saeki, M., Matsushita, K., Adachi, O., and Lidstrom, M. E. (2002) PqqC/D, which converts a biosynthetic intermediate to pyrroloquinoline quinone, *Biochem. Bioph. Res. Co. 299*, 268-272.

[48] Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004) Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7913-7918.

[49] Magnusson, O. T., Toyama, H., Saeki, M., Schwarzenbacher, R., and Klinman, J. P. (2004) The structure of a biosynthetic intermediate of pyrroloquinoline quinone (PQQ) and elucidation of the final step of PQQ biosynthesis, *J. Am. Chem. Soc. 126*, 5342-5343.

[50] Bonnot, F., Iavarone, A. T., and Klinman, J. P. (2013) Multistep, Eight-Electron Oxidation Catalyzed by the Cofactorless Oxidase, PqqC: Identification of

Chemical Intermediates and Their Dependence on Molecular Oxygen, *Biochemistry 52,* 4667-4675.

[51] Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Reed, J. C., and Liddington, R. C. (2004) Crystal structure of PqqC from Klebsiella pneumoniae at 2.1 A resolution, *Proteins 56*, 401-403.

[52] Puehringer, S., RoseFigura, J., Metlitzky, M., Toyama, H., Klinman, J. P., and Schwarzenbacher, R. (2010) Structural studies of mutant forms of the PQQ-forming enzyme PqqC in the presence of product and substrate, *Proteins 78*, 2554-2562.

[53] Khaliullin, B., Aggarwal, P., Bubas, M., Eaton, G. R., Eaton, S. S., and Latham, J. A. (2016) Mycofactocin biosynthesis: modification of the peptide MftA by the radical S-adenosylmethionine protein MftC, *FEBS Lett. 590*, 2538-2548.

[54] Bruender, N. A., and Bandarian, V. (2016) The Radical S-Adenosyl-l-methionine Enzyme MftC Catalyzes an Oxidative Decarboxylation of the C-Terminus of the MftA Peptide, *Biochemistry 55*, 2813-2816.

[55] Benjdia, A., Guillot, A., Lefranc, B., Vaudry, H., Leprince, J., and Berteau, O. (2016) Thioether bond formation by SPASM domain radical SAM enzymes: Calpha H-atom abstraction in subtilosin A biosynthesis, *Chem. Commun. (Camb.) 52*, 6249-6252.

[56] Fluhe, L., Knappe, T. A., Gattner, M. J., Schafer, A., Burghaus, O., Linne, U., and Marahiel, M. A. (2012) The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A, *Nat. Chem. Biol. 8*, 350-357.

[57] Bruender, N. A., and Bandarian, V. (2016) SkfB Abstracts a Hydrogen Atom from Calpha on SkfA To Initiate Thioether Cross-Link Formation, *Biochemistry 55*, 4131-4134.

[58] Fluhe, L., Burghaus, O., Wieckowski, B. M., Giessen, T. W., Linne, U., and Marahiel, M. A. (2013) Two [4Fe-4S] clusters containing radical SAM enzyme SkfB catalyze thioether bond formation during the maturation of the sporulation killing factor, *J. Am. Chem. Soc. 135*, 959-962.

[59] Wieckowski, B. M., Hegemann, J. D., Mielcarek, A., Boss, L., Burghaus, O., and Marahiel, M. A. (2015) The PqqD homologous domain of the radical SAM enzyme ThnB is required for thioether bond formation during thurincin H maturation, *FEBS Lett. 589*, 1802-1806.

[60] Bruender, N. A., Wilcoxen, J., Britt, R. D., and Bandarian, V. (2016) Biochemical and Spectroscopic Characterization of a Radical S-Adenosyl-L-methionine Enzyme Involved in the Formation of a Peptide Thioether Cross-Link, *Biochemistry 55*, 2122-2134.

[61] Shisler, K. A., and Broderick, J. B. (2012) Emerging themes in radical SAM chemistry, *Curr. Opin. Struct. Biol. 22*, 701-710.

[62] Goldman, P. J., Grove, T. L., Sites, L. A., McLaughlin, M. I., Booker, S. J., and Drennan, C. L. (2013) X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification, *Proc. Natl. Acad. Sci. U.S.A. 110*, 8519-8524.

[63] Tsai, T. Y., Yang, C. Y., Shih, H. L., Wang, A. H., and Chou, S. H. (2009) Xanthomonas campestris PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier, *Proteins 76*, 1042-1048.

[64] Rhodes (2006) *Crystallography Made Crystal Clear*, 3rd ed., Academic Press, Elsevier.

[65] Trueblood, K. N., Glusker, J.P. (2010) *Crystal Structure Analysis: A Primer*, 3rd ed., IUCR Oxford Science Publications, Oxford, UK.

[66] Dauter, Z., (Ed.) (2013) *SAD/MAD Phasing*, Springer, Dordrecht.

[67] Edwards, J. C. (2008) *Principles of NMR*, Process NMR Associates LLC, 87A Sand Pit Rd, Danbury CT 06810.

[68] Goosen, N., Horsman, H. P., Huinen, R. G., de Groot, A., and van de Putte, P. (1989) Genes involved in the biosynthesis of PQQ from Acinetobacter calcoaceticus, *Antonie Van Leeuwenhoek 56*, 85-91.

[69] Houck, D. R., Hanners, J. L., Unkefer, C. J., van Kleef, M. A., and Duine, J. A. (1989) PQQ: biosynthetic studies in Methylobacterium AM1 and Hyphomicrobium X using specific 13C labeling and NMR, *Antonie Van Leeuwenhoek 56*, 93-101.

[70] Unkefer, C. J., Houck, D. R., Britt, B. M., Sosnick, T. R., and Hanners, J. L. (1995) Biogenesis of pyrroloquinoline quinone from 3C-labeled tyrosine, *Methods Enzymol. 258*, 227-235.

[71] van Kleef, M. A., and Duine, J. A. (1988) L-tyrosine is the precursor of PQQ biosynthesis in Hyphomicrobium X, *FEBS Lett. 237*, 91-97.

[72] Anthony, C. (2001) Pyrroloquinoline Quinone (PQQ) and Quinoprotein Enzymes, *Antioxid. Redox Signal 3*, 757-774.

[73] Choi, O., Kim, J., Kim, J. G., Jeong, Y., Moon, J. S., Park, C. S., and Hwang, I. (2008) Pyrroloquinoline quinone is a plant growth promotion factor produced by Pseudomonas fluorescens B16, *Plant. Physiol. 146*, 657-668.

[74] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277-293.

[75] Live, D. H., Davis, D. G., Agosta, W. C., and Cowburn, D. (1984) Long-Range Hydrogen-Bond Mediated Effects in Peptides - N-15 Nmr-Study of Gramicidin-S in Water and Organic-Solvents, *J. Am. Chem. Soc. 106*, 1939-1941.

[76] Goddard, T. D. a. K., D.G. SPARKY 3 (software), University of California, San Francisco.

[77] Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., and Montelione, G. T. (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence, *J. Mol. Biol. 269*, 592-610.

[78] Muhandiram, D. R., and Kay, L. E. (1994) Gradient-Enhanced Triple-Resonance 3-Dimensional Nmr Experiments with Improved Sensitivity, *J. Magn. Reson. B 103*, 203-216.

[79] Kay, L. E., Xu, G.Y., Yamazaki, T. (1994) Enhanced-sensitivity triple- resonance spectroscopy with minimal H2O saturation, *J. Magn. Reson. A 109*, 129–133.

[80] Kay, L. E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., Forman-Kay, J.D. (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins, *J. Magn. Reson. B 101*, 333–337.

[81] Montelione, G. T., Lyons, B. A., Emerson, S. D., and Tashiro, M. (1992) An Efficient Triple Resonance Experiment Using C-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins, *J. Am. Chem. Soc. 114*, 10974-10975.

[82] Vuister, G. W., and Bax, A. (1994) Measurement of four-bond HN-H alpha J-couplings in staphylococcal nuclease, *J. Biomol. NMR 4*, 193-200.

[83] Yamazaki T., F.-K. J. D., Kay L.E. (1993) Two-dimensional NMR experiments for correlating 13Cb. and 1Hd/e chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings, *J. Am. Chem. Soc. 115*, 11054–11055.

[84] Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR 44*, 213-223.

[85] Wishart, D. S., and Sykes, B. D. (1994) The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data, *J. Biomol. NMR 4*, 171-180.

[86] Rucker, R., Chowanadisai, W., and Nakano, M. (2009) Potential Physiological Importance of Pyrroloquinoline Quinone, *Altern. Med. Rev. 14*, 268-277.

[87] Choi, O., Kim, J., Kim, J. G., Jeong, Y., Moon, J. S., Park, C. S., and Hwang, I. (2008) Pyrroloquinoline quinone is a plant growth promotion factor produced by Pseudomonas fluorescens B16, *Plant Physiol. 146*, 657-668.

[88] McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Nat. Prod. Rep. 26*, 537-559.

[89] Sanchez-Barrena, M. J., Martinez-Ripoll, M., Galvez, A., Valdivia, E., Maqueda, M., Cruz, V., and Albert, A. (2003) Structure of bacteriocin AS-48: from soluble state to membrane bound state, *J. Mol. Biol. 334*, 541-549.

[90] Westerling, J., Frank, J., and Duine, J. A. (1979) The prosthetic group of methanol dehydrogenase from Hyphomicrobium X: electron spin resonance evidence for a quinone structure, *Biochem. Biophys. Res. Commun. 87*, 719-724.

[91] Duine, J. A., and Frank, J., Jr. (1980) The prosthetic group of methanol dehydrogenase. Purification and some of its properties, *Biochem. J. 187*, 221-226.

[92] Haft, D. H. (2011) Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners, *BMC Genomics 12*, 21.

[93] Ameyama, M., Matsushita, K., Shinagawa, E., Hayashi, M., and Adachi, O. (1988) Pyrroloquinoline quinone: excretion by methylotrophs and growth stimulation for microorganisms, *Biofactors 1*, 51-53.

[94] Sode, K., Ito, K., Witarto, A. B., Watanabe, K., Yoshida, H., and Postma, P. (1996) Increased production of recombinant pyrroloquinoline quinone (PQQ) glucose

dehydrogenase by metabolically engineered Escherichia coli strain capable of PQQ biosynthesis, *J. Biotechnol. 49*, 239-243.

[95] Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nat. Chem. Biol. 6*, 9-18.

[96] Houck, D. R., Hanners, J. L., and Unkefer, C. J. (1988) Biosynthesis of Pyrroloquinoline Quinone .1. Identification of Biosynthetic Precursors Using C-13 Labeling and Nmr-Spectroscopy, *J. Am. Chem. Soc. 110*, 6920-6921.

[97] Houck, D. R., Hanners, J. L., and Unkefer, C. J. (1991) Biosynthesis of Pyrroloquinoline Quinone .2. Biosynthetic Assembly from Glutamate and Tyrosine, *J. Am. Chem. Soc. 113*, 3162-3166.

[98] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol. 7*.

[99] McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., and Lopez, R. (2013) Analysis Tool Web Services from the EMBL-EBI, *Nucleic Acids Res. 41*, W597-600.

[100] Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N., and Lopez, R. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework, *Nucleic Acids Res. 43*, W580-584.

[101] Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135-140.

[102] Evans, R. L., 3rd, Latham, J. A., Klinman, J. P., Wilmot, C. M., and Xia, Y. (2016) (1)H, (13)C, and (15)N resonance assignments and secondary structure information for Methylobacterium extorquens PqqD and the complex of PqqD with PqqA, *Biomol. NMR Assign. 10*, 385-389.

[103] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008) BioMagResBank, *Nucleic Acids Res. 36*, D402-408.

[104] Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006) Speeding up three-dimensional protein NMR experiments to a few minutes, *J. Am. Chem. Soc. 128*, 9042-9043.

[105] Gustavsson, M., Traaseth, N. J., Karim, C. B., Lockamy, E. L., Thomas, D. D., and Veglia, G. (2011) Lipid-mediated folding/unfolding of phospholamban as a regulatory mechanism for the sarcoplasmic reticulum Ca2+-ATPase, *J. Mol. Biol. 408*, 755-765.

[106] Gal, M., Kern, T., Schanda, P., Frydman, L., and Brutscher, B. (2009) An improved ultrafast 2D NMR experiment: towards atom-resolved real-time studies of protein kinetics at multi-Hz rates, *J. Biomol. NMR 43*, 1-10.

[107] Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289-302.

[108] Schwieters, C. D., Kuszewski, J. J., and Clore, G. M. (2006) Using Xplor-NIH for NMR molecular structure determination, *Prog. Nucl. Mag. Res. Sp. 48*, 47-62.

[109] Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson. 160*, 65-73.

[110] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. Sec. D 66*, 12-21.

[111] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-242.

[112] Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank, *Nat. Struct. Biol. 10*, 980.

[113] The PyMOL Molecular Graphics System, 1.8 ed., (molecular modeling software), Schrödinger, LLC.

[114] Wecksler, S. R., Stoll, S., Iavarone, A. T., Imsand, E. M., Tran, H., Britt, R. D., and Klinman, J. P. (2010) Interaction of PqqE and PqqD in the pyrroloquinoline

quinone (PQQ) biosynthetic pathway links PqqD to the radical SAM superfamily, *Chem. Commun. 46*, 7031-7033.

[115] Wishart, D. S., and Sykes, B. D. (1994) Chemical-Shifts as a Tool for Structure Determination, *Nuclear Magnetic Reson. Pt. C 239*, 363-392.

[116] Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A., and Nair, S. K. (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB, *Nature 517*, 509-512.

[117] Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella, *Proc. Natl. Acad. Sci. U.S.A. 102*, 7315-7320.

[118] McIntosh, J. A., Lin, Z., Tianero, M. D., and Schmidt, E. W. (2013) Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived Claisen rearrangements, *ACS Chem. Biol. 8*, 877-883.

[119] Regni, C. A., Roush, R. F., Miller, D. J., Nourse, A., Walsh, C. T., and Schulman, B. A. (2009) How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic, *EMBO J. 28*, 1953-1964.

[120] Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M., and Naismith, J. H. (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing, *Nat. Chem. Biol. 11*, 558-563.

[121] Kluskens, L. D., Kuipers, A., Rink, R., de Boef, E., Fekken, S., Driessen, A. J., Kuipers, O. P., and Moll, G. N. (2005) Post-translational modification of therapeutic peptides by NisB, the dehydratase of the lantibiotic nisin, *Biochemistry 44*, 12827-12834.

[122] Saichana, N., Tanizawa, K., Pechousek, J., Novak, P., Yakushi, T., Toyama, H., and Frebortova, J. (2016) PqqE from Methylobacterium extorquens AM1: a radical S-adenosyl-l-methionine enzyme with an unusual tolerance to oxygen, *J. Biochem. 159*, 87-99.

[123] Schramma, K. R., Bushin, L. B., and Seyedsayamdost, M. R. (2015) Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink, *Nat. Chem. 7*, 431-437.

[124] Azizi, A., Azizi, S., Heshmatian, B., and Amini, K. (2014) Improvement of functional recovery of transected peripheral nerve by means of chitosan grafts filled with vitamin E, pyrroloquinoline quinone and their combination, *Int. J. Surg. 12*, 76-82.

[125] Gong, D., Geng, C., Jiang, L., Aoki, Y., Nakano, M., and Zhong, L. (2012) Effect of pyrroloquinoline quinone on neuropathic pain following chronic constriction injury of the sciatic nerve in rats, *Eur. J. Pharmacol. 697*, 53-58.

[126] Zhang, L., Liu, J., Cheng, C., Yuan, Y., Yu, B., Shen, A., and Yan, M. (2012) The neuroprotective effect of pyrroloquinoline quinone on traumatic brain injury, *J. Neurotrauma 29*, 851-864.

[127] Zhang, Q., Ding, M., Cao, Z., Zhang, J., Ding, F., and Ke, K. (2013) Pyrroloquinoline quinine protects rat brain cortex against acute glutamate-induced neurotoxicity, *Neurochem. Res. 38*, 1661-1671.

[128] Davidson, V. L. (2008) Protein control of true, gated, and coupled electron transfer reactions, *Acc. Chem. Res. 41*, 730-738.

[129] Wang, Y., Li, J., and Liu, A. (2017) Oxygen activation by mononuclear nonheme iron dioxygenases involved in the degradation of aromatics, *J. Biol. Inorg. Chem. 22*, 395-405.

[130] Raven, E. L. (2017) A short history of heme dioxygenases: rise, fall and rise again, *J. Biol. Inorg. Chem. 22*, 175-183.

[131] Quist, D. A., Diaz, D. E., Liu, J. J., and Karlin, K. D. (2017) Activation of dioxygen by copper metalloproteins and insights from model complexes, *J. Biol. Inorg. Chem. 22*, 253-288.

[132] Peck, S. C., and van der Donk, W. A. (2017) Go it alone: four-electron oxidations by mononuclear non-heme iron enzymes, *J. Biol. Inorg. Chem. 22*, 381-394.

[133] Kal, S., and Que, L. (2017) Dioxygen activation by nonheme iron enzymes with the 2-His-1-carboxylate facial triad that generate high-valent oxoiron oxidants, *J. Biol. Inorg. Chem. 22*, 339-365.

[134] Huang, X., and Groves, J. T. (2017) Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C-H activation, *J. Biol. Inorg. Chem. 22*, 185-207.

[135] Fiedler, A. T., and Fischer, A. A. (2017) Oxygen activation by mononuclear Mn, Co, and Ni centers in biology and synthetic complexes, *J. Biol. Inorg. Chem. 22*, 407-424.

[136] Evans, R. L., 3rd, Latham, J. A., Xia, Y., Klinman, J. P., and Wilmot, C. M. (2017) Nuclear Magnetic Resonance Structure and Binding Studies of PqqD, a Chaperone Required in the Biosynthesis of the Bacterial Dehydrogenase Cofactor Pyrroloquinoline Quinone, *Biochemistry 56*, 2735-2746.

[137] Stepanov, S., Makarov, O., Hilgart, M., Pothineni, S. B., Urakhchin, A., Devarapalli, S., Yoder, D., Becker, M., Ogata, C., Sanishvili, R., Venugopalan, N., Smith, J. L., and Fischetti, R. F. (2011) JBluIce-EPICS control system for macromolecular crystallography, *Acta Crystallogr. Sec. D 67*, 176-188.

[138] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr. Sec. D 66*, 133-144.

[139] Kabsch, W. (2010) XDS, *Acta Crystallogr. Sec. D 66*, 125-132.

[140] Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement, *J. Appl. Crystallogr 41*, 641-643.

[141] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr 40*, 658-674.

[142] McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser, *Acta Crystallogr. Sec. D 63*, 32-41.

[143] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr. Sec. D 67*, 235-242.

[144] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. Sec. D 60*, 2126-2132.

[145] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res. 35*, W375-383.

[146] Ten Eyck, L. F. (1973) Crystallographic fast Fourier transforms, *Acta Crystallogr. Sec. A 29*, 8.

[147] Magnusson, O. T., RoseFigura, J. M., Toyama, H., Schwarzenbacher, R., and Klinman, J. P. (2007) Pyrroloquinoline quinone biogenesis: characterization of PqqC and its H84N and H84A active site variants, *Biochemistry 46*, 7174-7186.

[148] Fisher, S. J., Puehringer, S. (2014) Title given for deposition of 4NY7: Bond length analysis of the PqqC Y175F mutant structure shows evidence for bound PQQ in the reduced form, *unpublished*.

[149] Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding, *Prog. Nucl. Mag. Res. Sp. 73*, 1-16.

[150] Murthy, B. N., Sinha, S., Surolia, A., Indi, S. S., and Jayaraman, N. (2008) SPR and ITC determination of the kinetics and the thermodynamics of bivalent versus monovalent sugar ligand-lectin interactions, *Glycoconj. J. 25*, 313-321.

[151] Johnson, B. J., Yukl, E. T., Klema, V. J., Klinman, J. P., and Wilmot, C. M. (2013) Structural snapshots from the oxidative half-reaction of a copper amine oxidase: implications for O2 activation, *J. Biol. Chem. 288*, 28409-28417.

[152] De la Mora-Rey, T., and Wilmot, C. M. (2007) Synergy within structural biology of single crystal optical spectroscopy and X-ray crystallography, *Curr. Opin. Struct. Biol. 17*, 580-586.

[153] Wilmot, C. M., Sjogren, T., Carlsson, G. H., Berglund, G. I., and Hajdu, J. (2002) Defining redox state of X-ray crystal structures by single-crystal ultraviolet-visible microspectrophotometry, *Methods Enzymol. 353*, 301-318.